Out-of-sample equity premium prediction: A complete subset quantile regression approach

Loukia Meligkotsidou[a], Ekaterini Panopoulou[b]*, Ioannis D.Vrontos[c], Spyridon D. Vrontos[d]

[a]Department of Mathematics, University of Athens, Greece

[b]Kent Business School, University of Kent, Canterbury, United Kingdom

[c]Department of Statistics, Athens University of Economics and Business, Greece

[d]Department of Mathematical Sciences, University of Essex, United Kingdom

Abstract

This paper extends the complete subset linear regression framework to a quantile regression setting. We employ complete subset combinations of quantile forecasts in order to construct robust and accurate equity premium predictions. Our recursive algorithm that selects, in real time, the best complete subset for each predictive regression quantile succeeds in identifying the best subset in a time- and quantile-varying manner. We show that our approach delivers statistically and economically significant out-of-sample forecasts relative to both the historical average benchmark and the complete subset mean regression approach.

JEL classification: G11; G12; C22; C53

Keywords: Equity premium; Forecast combination; Predictive quantile regression; Robust point forecasts; Subset quantile regressions.

*Corresponding author. Ekaterini Panopoulou, Kent Business School, University of Kent, Canterbury CT2 7PE, United Kingdom, Tel.: 0044 1227 824469, Email: A.Panopoulou@kent.ac.uk.

1 Introduction

The issue of forecasting equity returns is one of the most widely discussed topics in the finance literature mainly due to its central role in asset pricing, portfolio allocation and evaluation of investment managers. The in-sample predictive ability of a quite exhaustive list of potential predictors that typically contains valuation ratios, various interest rates and spreads, distress indicators, inflation rates along with other macroeconomic variables, indicators of corporate activity, etc. was the focus of the earlier studies.[1] However, since the seminal contribution of Goyal and Welch (2008) who show that their long list of predictors can not deliver consistently superior out-of-sample performance, attention has turned to the development of improved forecasting methods in order to establish the empirical validity of equity premium predictability.[2] To mention a few, Campbell and Thompson (2008) show that when imposing simple restrictions, suggested by economic theory, on predictive regressions' coefficients, the out-of-sample performance improves. Based on their result, the authors argue that market timing strategies can deliver profits to investors (see also Ferreira and Santa-Clara (2011)). Ludvigson and Ng (2007) and Neely, Rapach, Tu and Zhou (2014) adopt a diffusion index approach, which can conveniently track the key movements in a large set of predictors, and they find evidence of improved equity premium forecasting ability.[3]

In an attempt to reduce both model uncertainty and parameter instability, Rapach, Strauss and Zhou (2010) employ forecast combinations of univariate equity premium models and find that combinations of individual single variable predictive regression models significantly beat the historical average forecast. Building on Rapach, Strauss and Zhou (2010), Meligkotsidou, Panopoulou, Vrontos and Vrontos (2014, MPVV henceforth) incorporate the forecast combination methodology in a quantile regression setting. Their quantile regression approach to equity premium prediction allows them to cope with the non-linearity and non-normality patterns that are evident in the relationship between stock returns and potential predictors. In this way,

[1] Commonly used valuation ratios are the dividend price/dividend yield ratio (see for example, Fama and French, 1988, 1989), the earnings price ratio (Campbell and Shiller, 1988, 1998), and the book-to-market ratio (Kothari and Shanken, 1997). Another strand of the literature includes macroeconomic/ financial variables such as inflation rates, short-term and long-term interest rates along with term and corporate bond spreads in the set of predictors (see e.g. Fama and Schwert, 1977; Campbell and Vuolteenaho, 2004; Campbell, 1987; Fama and French, 1989; Ang and Bekaert, 2007). Lettau and Ludvigson (2001) finds that the consumption to wealth ratio helps equity premium predictability, while corporate financing activity is exploited in Baker and Wurgler (2000). A comprehensive list of variables that serve as predictors can be found in Goyal and Welch (2008).

[2] Following the related literature, equity premium is proxied by excess returns.

[3] Rapach and Zhou (2012) offer a detailed review on the issue of equity return predictability.

robust and accurate equity premium forecasts are produced by combining a set of predictive quantile regressions in either a fixed or time-varying manner. A novel forecast combination method based on complete subset regressions is put forward by Elliott, Gargano and Timmermann (2013, EGT henceforth). The authors propose combining forecasts from all possible linear regression models that keep the number of predictors fixed. Their empirical application on equity premium predictability shows that subset combinations of up to four predictors generates superior forecast accuracy.

This paper proposes a new forecasting approach based on complete subset quantile regressions. Specifically, we extend the framework of EGT to a quantile regression setting and adopt the methodology of MPVV to this subset quantile regression framework, in order to produce robust and accurate equity premium forecasts. Our proposed methodology merges three strands of the literature on out-of-sample forecasting and, as shown, exploits the benefits emerging from each one. First, we exploit the ability of the quantile regression setting to produce robust and accurate point forecasts. Second, we reduce model uncertainty and parameter instability by employing quantile forecast combinations. Finally, we employ complete subset quantile regressions which induces shrinkage to the respective estimates and further helps reduce the effect of parameter estimation error.

To be more specific, our forecasting framework is rooted in quantile predictive regressions, which have attracted a vast amount of attention since the seminal paper of Koenker and Bassett (1978).[4] Empirical contributions in the field of finance include Bassett and Chen (2001), Engle and Manganelli (2004), Meligkotsidou, Vrontos and Vrontos (2009), Cenesizoglou and Timmermann (2012), Chuang, Kuan and Lin (2009) and Baur, Dimpfl and Jung (2012). The main advantage of our quantile regression framework lies in its ability to cope with non-linearity and non-normality patterns in the joint relationship between equity returns and candidate predictors (see, inter alia, Guidolin and Timmermann, 2009; Guidolin, Hyde, McMillan and Ono, 2009; Henkel, Martin and Nadari, 2011). Robust point forecasts of the equity premium or any variable of interest in general can be constructed as weighted averages of a set of quantile forecasts by employing either fixed weighting or time-varying weighting schemes.

Incorporating the forecast combination approach (see Rapach, Strauss and Zhou, 2010) into our quantile regression setting helps reduce model uncertainty and deals with parameter insta-

[4] See also Buchinsky (1994, 1995) and Yu, Lu and Stander (2003).

bility.[5] MPVV propose two alternative ways to generate forecasts within the quantile regression setup. The first approach proceeds by first constructing robust point forecasts from a set of quantile predictions all of which are based on the same predictive variable. Next, it combines the robust forecasts obtained from different predictors using several existing combination methods in order to produce a final point forecast. The second approach consists of first combining all the predictions of the same quantile obtained from different single predictor model specifications, in order to produce combined quantile forecasts. Then, robust point forecasts are obtained by operating either a fixed or a time-varying weighting scheme on the combined quantile forecasts.

The methodologies discussed so far employ single variable models in either a linear or a quantile regression framework. EGT abstract from the single predictor models and propose combining forecasts from all possible linear regression models that keep the number of predictors fixed. Their approach introduces a complex version of shrinkage to the respective estimates which helps reduce the effect of parameter estimation error.[6] EGT show that the amount of shrinkage induced on least squares estimates from subset regressions is a function of the number of variables included in the model (k) and the total number of available predictors (K). Given that the amount of shrinkage depends on all the least squares estimates, it varies with each coefficient. Moreover, this methodology can cure the omitted variable bias especially in cases with strongly positively correlated regressors. The authors propose constructing forecasts based on a simple averaging scheme of all the possible models employed keeping the numbers of regressors fixed. In this paper, we extend the framework of EGT to the quantile predictive regression framework discussed above. Similarly to EGT, we utilize information from all the predictors simultaneously in order to produce combined quantile forecasts from all quantile regressions that keep the number of predictors fixed. We also abstract from the simple averaging schemes and introduce several existing combination schemes into our setting. Then, the obtained quantile forecasts are synthesized to produce robust point forecasts of the variable of interest.

The empirical findings of both EGT and the present paper suggest that the predictive performance of subset regressions highly depend on the value of k. A further contribution of this paper is the development of a recursive algorithm for selecting k in real time, based on the

[5] Timmermann (2006) provides a detailed review on forecast combination methodologies.

[6] Shrinkage typically is employed in order to limit the number of parameters that have to be estimated when many potential predictors are available. Contributions to this field include the ridge regression (Hoerl and Kennard, 1970), model averaging (Bates and Granger, 1969; Raftery, Madigan and Hoeting, 1997), bagging (Breiman, 1996) and the Lasso (Tibshirani, 1996).

past history of excess returns and predictive variables. The proposed algorithm is a likelihood-based method that chooses the best complete subset for a given quantile and is flexible enough to allow for variability of the selected value of k across quantiles. In this way, our approach incorporates information on the best subset for each quantile of the return distribution in real time and these 'optimal' quantile forecasts are appropriately combined to deliver robust equity premium forecasts.

To anticipate our key results, we find that our complete subset quantile regression framework achieves superior predictive performance, both in statistical and economic evaluation terms. More in detail, our proposed approach can lead to an out-of-sample R^2 of 5.71% (relative to the historical average benchmark) as opposed to 4.10% of the subset linear regression approach of EGT and 3.58% of the combination approach of Rapach, Strauss and Zhou (2010). While in a linear regression framework, subsets of two variables ($k = 2$) perform better than the remaining specifications, in our quantile regression framework subsets of three variables ($k = 3$) emerge as superior. More importantly, our real time recursive algorithm for selecting k across quantiles of returns succeeds in identifying the 'correct' value of k which is both time-varying and quantile-varying. When evaluating our forecasts from an economic perspective and specifically for a mean-variance investor, we also need return volatility forecasts, which we construct using the interval approximation approach of Pearson and Tukey (1965) and a set of predictive quantiles. Our economic evaluation results suggest that an investor that adopts our framework can gain sizable benefits which range from 3.91% to an impressive 6.27% per year relative to a naive strategy based on the historical benchmark performance.

The outline of the paper is as follows. Section 2 describes the complete subset regression framework of EGT and introduces its extension to the quantile regression framework. The proposed methodology for robust estimation of the central location of the distribution of returns is outlined in Section 3. Section 4 presents our empirical findings, while section 5 describes the proposed methodology for the recursive selection of the number of predictors. Section 6 outlines the economic evaluation framework and presents the associated findings. Section 7 summarizes and concludes.

2 Complete Subset Quantile Regressions

In this section we present the setup for our analysis. Section 2.1 outlines the EGT complete subset regressions framework and Section 2.2 extends this framework to subset quantile regressions. Section 2.3 proposes two novel forecasting approaches based on complete subset quantile regressions.

2.1 Complete subset regressions

EGT propose a new method for combining forecasts based on complete subset regressions. For a given set of potential predictors, the authors propose combining forecasts from all possible linear regressions that keep the number of predictors fixed. For K possible predictors, there are K univariate models and $n_{k,K} = K!/((K-k)!k!)$ different $k-$variate models for $k \leq K$. The set of models for a fixed value of k is referred to as a complete subset and the authors propose using equal-weighted combinations of the forecasts from all models within these subsets indexed by k.

More in detail, suppose that we are interested in forecasting the equity premium, denoted by r_t, using a set of K predictive variables. First we consider all possible predictive mean regression models with a single predictor, i.e. $k = 1$, of the form

$$r_{t+1} = \alpha_i + \beta_i x_{it} + \varepsilon_{t+1},\ i = 1, \ldots, K, \tag{1}$$

where r_{t+1} is the observed excess return on a stock market index in excess of the risk-free interest rate at time $t+1$, x_{it} are the K observed predictors at time t, and the error terms ε_{t+1} are assumed to be independent with mean zero and variance σ^2. The predictive mean regression models can be estimated using the Ordinary Least Squares (OLS) method by minimizing the sample estimate of the quadratic expected loss, $\sum_{t=0}^{T-1} (r_{t+1} - \alpha_i - \beta_i x_{it})^2$, or the Maximum Likelihood (ML) approach after specifying the parametric form of the error distribution[7]. Similarly, a regression of r_{t+1} can be run on a particular subset of the regressors and then average the forecasts across all k dimensional subsets to provide the forecast for the variable of interest, where $k \leq K$. EGT show that while subset regression combinations bear similarities

[7] The sample size T denotes any estimation sample employed in our recursive forecasting experiment. Details on the forecasting design are given in Section 4.

to a complex version of shrinkage, they do not reduce to shrinking OLS estimates. Rather the coefficient that controls shrinkage depends on all OLS estimates, the dimension of the subset and the number of included predictors. Only in the case of orthonormal regressors does subset regression reduce to ridge regression. Moreover, the amount of shrinkage imposed on each coefficient differs with the coefficient at hand. More importantly, the authors show that in the case of strongly correlated predictors, subset regression can remedy the omitted variable bias and improve forecasts. While the authors use equal-weighted combinations of forecasts within each subset along with approximate Bayesian Model Averaging, alternative weighting schemes can be employed. To this end, we also employ the Median, the Trimmed Mean, the Discount Mean Squared Forecast Error (DMSFE) of Stock and Watson (2004) along with the Cluster combining method, introduced by Aiolfi and Timmermann (2006).[8]

2.2 Complete subset quantile regressions

The above linear subset regression specification can only predict the mean and not the entire distribution of returns in the event that the joint distribution of r_{t+1} and x_{it} is not bivariate Gaussian and, therefore, their relationship is not linear. Following the literature on the non-linear relationship between returns and predictors (Guidolin and Timmermann, 2009; Guidolin, Hyde, McMillan and Ono, 2009; Chen and Hong, 2010; Henkel, Martin and Nadari, 2011) we adopt a more sophisticated approach to equity premium forecasting by employing predictive quantile regression models (Koenker and Bassett, 1978; Buchinsky, 1998; Yu, Lu and Stander, 2003). In this paper we incorporate the complete subset combination framework of EGT in our quantile regression setting. The proposed approach is designed as follows.

First, consider single predictor quantile regression models ($k = 1$) of the form

$$r_{t+1} = \alpha_i^{(\tau)} + \beta_i^{(\tau)} x_{it} + \varepsilon_{t+1},\ i = 1, \ldots, K, \tag{2}$$

where $\tau \in (0, 1)$ and the errors ε_{t+1} are assumed independent from an error distribution $g_\tau(\varepsilon)$ with the τth quantile equal to 0, i.e. $\int_{-\infty}^{0} g_\tau(\varepsilon) d\varepsilon = \tau$. Model (2) suggests that the τth quantile of r_{t+1} given x_{it} is $Q_\tau(r_{t+1}|x_{it}) = \alpha_i^{(\tau)} + \beta_i^{(\tau)} x_{it}$, where the intercept and the regression

[8] To keep the analysis clear, Appendix A.1 provides a detailed description of the formation of these weighting schemes.

coefficients depend on τ. The $\beta_i^{(\tau)}$'s are likely to vary across τ's, revealing a larger amount of information about returns than the predictive mean regression model (Equation 1). Estimators of the parameters of the linear quantile regression models in (2), $\hat{\alpha}_i{}^{(\tau)}, \hat{\beta}_i^{(\tau)}$, can be obtained by minimizing the sum $\sum_{t=0}^{T-1} \rho_\tau \left(r_{t+1} - \alpha_i{}^{(\tau)} - \beta_i^{(\tau)} x_{it} \right)$, where $\rho_\tau(u)$ is the asymmetric linear loss function, usually referred to as the check function,

$$\rho_\tau(u) = u\left(\tau - I(u < 0)\right) = \frac{1}{2}\left[|u| + (2\tau - 1)u\right]. \tag{3}$$

In the symmetric case of the absolute loss function ($\tau = 1/2$) we obtain estimators of the median predictive regression models. A parametric approach to inference on the quantile regression parameters arises if the error distribution $g_p(\varepsilon)$ is specified. The error distribution that has been widely used for parametric inference in the quantile regression literature is the asymmetric Laplace distribution (for details, see Yu and Moyeed, 2001, and Yu and Zhang, 2005) with probability density function

$$g_\tau(\varepsilon) = \frac{\tau(1-\tau)}{\sigma^{(\tau)}} \exp\left[-\frac{|\varepsilon| + (2\tau - 1)\varepsilon}{2\sigma^{(\tau)}} \right], 0 < \tau < 1, \sigma^{(\tau)} > 0. \tag{4}$$

For $\tau = 1/2$, corresponding to the median regression, (4) becomes the symmetric Laplace density. A likelihood function can be formed by combining T independent asymmetric Laplace densities of the form (4), i.e.

$$L^{(\tau)}\left(r_{1:T} | \alpha_i{}^{(\tau)}, \beta_i^{(\tau)}, \sigma^{(\tau)} \right) = \left(\frac{\tau(1-\tau)}{\sigma^{(\tau)}} \right)^T \exp\left\{ -\frac{1}{\sigma^{(\tau)}} \sum_{t=0}^{T-1} \rho_\tau \left(r_{t+1} - \alpha_i{}^{(\tau)} - \beta_i^{(\tau)} x_{it} \right) \right\}. \tag{5}$$

Then (5) can be used for likelihood based inference for the parameters $\alpha_i{}^{(\tau)}, \beta_i^{(\tau)}, \sigma^{(\tau)}$, for example for maximum likelihood estimation. The maximization of this likelihood function with respect to $\alpha_i{}^{(\tau)}, \beta_i^{(\tau)}$ is equivalent to minimizing the expected asymmetric linear loss, while the ML estimator of $\sigma^{(\tau)}$ is $\widehat{\sigma}^{(\tau)} = \frac{1}{T} \sum_{t=0}^{T-1} \rho_\tau \left(r_{t+1} - \alpha^{(\tau)} - \beta_i^{(\tau)} x_{it} \right)$. Similarly to the predictive mean regression case, the quantile regression (Equation 2) of r_{t+1} can be run on a particular subset (k) of the regressors K, $k \leq K$, with the aim to produce quantile forecasts of the equity premium. Then, we employ one of the approaches outlined in the next subsection in order to get robust and accurate point forecasts of the variable of interest.

The advantage of the parametric approach to inference is that it enables us to compare different quantile regression models, corresponding to different subsets of predictors, using criteria based on the likelihood function, for example the Bayesian Information Criterion (BIC) or Bayesian model comparison. This further enables us to establish an approach of selecting the best (in a likelihood based sense) complete subset on the basis of which forecasts are formed (see Section 5).

2.3 Forecasting Approaches based on Complete Subset Quantile Regression

This subsection outlines the two novel forecasting approaches we put forward. As already mentioned, these are based on subset quantile regressions and aim at producing robust and accurate point forecasts of the equity premium by taking advantage of the subset framework, the quantile regression framework and the information content in individual (or combined) potential predictors. Specifically, we construct equity premium point forecasts by combining quantile forecasts obtained from a set of complete subset regressions ($k - variate$ models with $k \leq K$). For each k, $n_{k,K}$ regressions are run in order to predict the τ^{th} quantile of the distribution of the next period's excess return (r_{t+1}). Next, two approaches are explored in order to combine these quantile forecasts into a point forecast that is robust to non-normality and non-linearity.

The first approach, which we name Robust Forecast Combination approach (RFC) proceeds by first combining the quantile forecasts across all values of τ into point forecasts for each complete subset of predictors. As outlined in the next section, we employ Tukey's (1977) and Gastwirth's (1966) three-quantile estimators and the five-quantile estimator of Judge, Hill, Griffiths, Lutkepohl and Lee (1988) along with their time-varying counterparts developed in MPVV. This step yields $n_{k,K}$ point forecasts which are further combined in order to reduce uncertainty risk associated with each subset of the predictive variables. Except for the simple averaging scheme, suggested by EGT, we also employ the Trimmed Mean, the Median, the Discount Mean Squared Forecast Error (DMSFE) of Stock and Watson (2004) along with the Cluster combining method, introduced by Aiolfi and Timmermann (2006). These combining schemes utilize the Mean Squared Forecast Error (MSFE) as a loss function. (Details are given in Appendix A.1).

The second approach, which we name Quantile Forecast Combination (QFC) consists of first combining the predicted τ^{th} quantiles across all different subsets (k) of predictors ($n_{k,K}$ model specifications). With the exception of the Mean, Trimmed Mean and Median combining methods, the existing combination methods are not appropriate for combining predictor information in the quantile regression context. To this end, the MSFE loss function has to be replaced by a metric based on the asymmetric linear loss function (Equation 3). Following MPVV, we employ the Discount Asymmetric Loss Forecast Error (DALFE) and the Asymmetric Loss Cluster (AL Cluster) in order to construct subset quantile forecasts (see Appendix A.2). This step yields a set of quantile forecasts (one for each τ_j), which are then combined into final robust point forecasts using either a fixed or a time-varying weighting scheme (see next section).

3 Robust Point Forecasts based on Regression Quantiles

In this section we consider the problem of constructing robust point forecasts of the equity premium based on a set of predictive quantile regressions as an alternative to the standard approach which produces forecasts based on the predictive mean regression model. Robust point estimates of the central location of a distribution can be constructed as weighted averages of a set of quantile estimators employing either fixed or time-varying weighting schemes.

3.1 Point Forecasts based on a Fixed Weighting Scheme

For a given model specification or a given complete subset that has been used for producing quantile forecasts, robust point forecasts can be constructed as weighted averages of a set of quantile forecasts. First, we employ standard estimators with fixed, prespecified weights of the form

$$\hat{r}_{t+1} = \sum_{\tau \in S} p_\tau \hat{r}_{t+1}(\tau), \quad \sum_{\tau \in S} p_\tau = 1,$$

where S denotes the set of quantiles that are combined, $\hat{r}_{t+1}(\tau)$ denotes the quantile forecasts associated with the τth quantile and $\hat{r}_{t+1}$ is the produced robust point forecast. Here the weights represent probabilities attached to different quantile forecasts, suggesting how likely to predict the return at the next period each regression quantile is.

We consider Tukey's (1977) trimean and the Gastwirth (1966) three-quantile estimator given, respectively, by the following formulae

$$\text{FW1:} \qquad \widehat{r}_{t+1} = 0.25\hat{r}_{t+1}(0.25) + 0.50\hat{r}_{t+1}(0.50) + 0.25\hat{r}_{t+1}(0.75)$$

$$\text{FW2:} \qquad \widehat{r}_{t+1} = 0.30\hat{r}_{t+1}(1/3) + 0.40\hat{r}_{t+1}(0.50) + 0.30\hat{r}_{t+1}(2/3).$$

In order to attach more weight on extreme positive and negative events, we also use the five-quantile estimator, suggested by Judge, Hill, Griffiths, Lutkepohl and Lee (1988).

$$\text{FW3: } \widehat{r}_{t+1} = 0.05\hat{r}_{t+1}(0.10) + 0.25\hat{r}_{t+1}(0.25) + 0.40\hat{r}_{t+1}(0.50) + 0.25\hat{r}_{t+1}(0.75) + 0.05\hat{r}_{t+1}(0.90).$$

3.2 Point Forecasts based on a Time-varying Weighting Scheme

Relaxing the assumption of a constant weighting scheme seems to be a natural extension. A number of factors, such as changes in regulatory conditions, market sentiment, monetary policies, institutional framework or even changes in macroeconomic interrelations (Campbell and Cochrane, 1999; Menzly, Santos and Veronesi, 2004; Dangl and Halling, 2012) can motivate the employment of time-varying schemes in the generation of robust point forecasts. Time-varying weighting schemes aim at producing an empirical model that allows for economic changes over time and is capable of determining the 'right' parameter values in time to help investors (Spiegel, 2008).

The variable of interest, r_{t+1}, is predicted using an optimal linear combination $\mathbf{p}_t=[p_{\tau,t}]_{\tau\in S}$ of the quantile forecasts $\hat{r}_{t+1}(\tau)$ given by

$$\hat{r}_{t+1} = \sum_{\tau\in S} p_{\tau,t}\hat{r}_{t+1}(\tau), \ \ \sum_{\tau\in S} p_{\tau,t} = 1.$$

The weights, $\mathbf{p}_t$, are estimated recursively using a holdout out-of-sample period continuously updated by one observation at each step. Optimal estimates of the weights are obtained by minimizing the mean squared forecast errors, $E_t(r_{t+1} - \hat{r}_{t+1})^2$, under an appropriate set of constraints. Our optimization procedure is the analogue of the constrained Granger and Ramanathan (1984) method for quantile regression forecasts (see also Timmermann, 2006; Hansen, 2008; Hsiao and Wan, 2014). Specifically, we employ constrained least squares using the quantile forecasts as regressors in lieu of a standard set of predictors. The time-varying weights on the quantile forecasts bear an interesting relationship to the portfolio weight constraints in finance. In this sense we constrain the weights to be non-negative, sum to one and not to exceed certain lower and upper bounds in order to reduce the weights' volatility and stabilize forecasts. In our empirical application, we employ three time-varying specifications which may be viewed as the time-varying counterparts of our FW1-FW3 schemes. More specifically, FW1 with time-varying coefficients becomes

$$\text{TVW1: } \widehat{r}_{t+1} = p_{0.25,t}\hat{r}_{t+1}(0.25) + p_{0.50,t}\hat{r}_{t+1}(0.50) + p_{0.75,t}\hat{r}_{t+1}(0.75),$$

where $p_{\tau,t}, \tau \in S = \{0.25, 0.50, 0.75\}$ are estimated by the optimization procedure

$$\begin{gathered}\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E[r_{t+1} - (p_{0.25,t}\hat{r}_{t+1}(0.25) + p_{0.50,t}\hat{r}_{t+1}(0.50) + p_{0.75,t}\hat{r}_{t+1}(0.75))]^2 \\ s.t.\ p_{0.25,t} + p_{0.50,t} + p_{0.75,t} = 1, 0.20 \leq p_{0.25,t} \leq 0.40, \\ 0.40 \leq p_{0.50,t} \leq 0.60, 0.20 \leq p_{0.75,t} \leq 0.40.\end{gathered}$$

Similarly, the FW2 scheme with time-varying coefficients becomes

$$\text{TVW2: } \widehat{r}_{t+1} = p_{1/3,t}\hat{r}_{t+1}(1/3) + p_{0.5,t}\hat{r}_{t+1}(0.50) + p_{2/3,t}\hat{r}_{t+1}(2/3),$$

where $p_{\tau,t}, \tau \in S = \{1/3, 0.50, 2/3\}$ are estimated by the following optimization procedure

$$\begin{gathered}\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E[r_{t+1} - (p_{1/3,t}\hat{r}_{t+1}(1/3) + p_{0.5,t}\hat{r}_{t+1}(0.50) + p_{2/3,t}\hat{r}_{t+1}(2/3))]^2 \\ s.t.\ p_{1/3,t} + p_{0.50,t} + p_{2/3,t} = 1, 0.15 \leq p_{1/3,t} \leq 0.45, \\ 0.30 \leq p_{0.5,t} \leq 0.50, 0.15 \leq p_{2/3,t} \leq 0.45.\end{gathered}$$

Finally, the FW3 scheme with time-varying coefficients becomes

$$\begin{aligned}\text{TVW3: } \widehat{r}_{t+1} &= p_{0.10,t}\hat{r}_{t+1}(0.10) + p_{0.25,t}\hat{r}_{t+1}(0.25) + p_{0.5,t}\hat{r}_{t+1}(0.50) \\ &+ p_{0.75,t}\hat{r}_{t+1}(0.75) + p_{0.90,t}\hat{r}_{t+1}(0.90),\end{aligned}$$

where $p_{\tau,t}, \tau \in S = \{0.10, 0.25, 0.50, 0.75, 0.90\}$ are estimated by the following optimization procedure

$$\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E[r_{t+1} - (p_{0.10,t}\hat{r}_{t+1}(0.10) + p_{0.25,t}\hat{r}_{t+1}(0.25)+$$

$$+p_{0.5,t}\hat{r}_{t+1}(0.5) + p_{0.75,t}\hat{r}_{t+1}(0.75) + p_{0.90,t}\hat{r}_{t+1}(0.90))]^2$$

$$s.t.\ p_{0.10,t} + p_{0.25,t} + p_{0.50,t} + p_{0.75,t} + p_{0.90,t} = 1$$

$$0.00 \leq p_{0.10,t} \leq 0.10, 0.15 \leq p_{0.25,t} \leq 0.35,$$

$$0.40 \leq p_{0.50,t} \leq 0.60, 0.15 \leq p_{0.75,t} \leq 0.35, 0.00 \leq, p_{0.90,t} \leq 0.10.$$

4 Empirical findings

4.1 Data, forecast construction and forecast evaluation

The data we employ are from Goyal and Welch (2008) who provide a detailed description of transformations and datasources.[9] The equity premium is calculated as the difference of the continuously compounded S&P500 returns, including dividends, and the Treasury Bill rate. Following the line of work of Goyal and Welch (2008), Rapach, Strauss and Zhou (2010) and Ferreira and Santa-Clara (2011), out-of-sample forecasts of the equity premium are generated by continuously updating the estimation window, i.e. following a recursive (expanding) window. More specifically, we divide the total sample of T observations into an in-sample portion of the first T_0 observations and an out-of-sample portion of $P = T - T_0$ observations used for forecasting. The estimation window is continuously updated following a recursive scheme, by adding one observation to the estimation sample at each step. As such, the coefficients in any predictive model employed are re-estimated after each step of the recursion. Proceeding in this way through the end of the out-of-sample period, we generate a series of P out-of-sample forecasts for the equity premium $\{\hat{r}_{i,t+1}\}_{t=T_0}^{T-1}$. Our forecasting experiment is conducted on a quarterly basis and data span 1947:1 to 2010:4. Our out-of-sample forecast evaluation period corresponds to the 'long' one analyzed by Goyal and Welch (2008) and Rapach, Strauss and Zhou (2010) covering the period 1965:1-2010:4.[10]

[9] The data are available at http://www.hec.unil.ch/agoyal/. We thank Prof. Goyal for making them available to us.

[10] Please note that the out-of-sample period refers to the period used to evaluate the out-of-sample forecasts. We use the ten years 1955:1 to 1964:4 (40 quarters) before the start of the out-of-sample evaluation period as the initial holdout out-of-sample period, required for both constructing our time-varying robust forecasts and for several forecast combination schemes.

The 12 economic variables employed in our analysis are related to stock-market characteristics, interest rates and broad macroeconomic indicators. With respect to stock market characteristics, we employ the Dividend–price ratio (log), D/P, the difference between the log of dividends paid on the S&P 500 index and the log of stock prices (S&P 500 index), where dividends are measured using a one-year moving sum; Dividend yield (log), D/Y, the difference between the log of dividends and the log of lagged stock prices; Earnings–price ratio (log), E/P, the difference between the log of earnings on the S&P 500 index and the log of stock prices, where earnings are measured using a one-year moving sum; Book-to-market ratio, B/M, the ratio of book value to market value for the Dow Jones Industrial Average and Net equity expansion, NTIS, the ratio of twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks. Turning to interest-rate related variables, we employ five variables ranging from short-term government rates to long-term government and corporate bond yields and returns along with their spreads. These are the Treasury bill rate, TBL, the interest rate on a three-month Treasury bill (secondary market); Long-term return, LTR, the return on long-term government bonds; Term spread, TMS, the difference between the long-term yield and the Treasury bill rate; Default yield spread, DFY, the difference between BAA- and AAA-rated corporate bond yields; Default return spread, DFR, the difference between long-term corporate bond and long-term government bond returns; To capture the overall macroeconomic environment, we employ the inflation rate, INFL, calculated from the CPI (all urban consumers) and the investment-to-capital ratio, I/K, the ratio of aggregate (private nonresidential fixed) investment to aggregate capital for the entire economy.[11]

The natural benchmark forecasting model is the historical mean or prevailing mean (PM) model, according to which the forecast of the equity premium coincides with the constant in the linear regression model (1) when no predictor is included, i.e. $k = 0$. As a measure of forecast accuracy, we employ the out-of-sample R^2 computed as $R^2_{OS} = 1 - \frac{MSFE_i}{MSFE_{PM}}$, where $MSFE_i$ is the Mean Square Forecast Error associated with each of our competing models and specifications and $MSFE_{PM}$ is the respective value for the PM model, both computed over the out-of-sample period. Positive values are associated with superior forecasting ability of our proposed model/specification. Given that point estimates of the R^2_{OS} are sample dependent, we

[11]Following EGT, we exclude the log dividend earnings ratio and the long term yield in order to avoid multicollinearity.

need to evaluate the statistical significance of our forecasts. To this end, we employ the Clark and West (2007) (CW) approximate normal test to compare our models/ specifications.[12]

The following subsections present an illustration of our proposed complete subset quantile regression approach to equity premium forecasting. The aim of our analysis is to assess the predictive ability of the proposed forecasting approaches and to compare their performance against that of alternative approaches used in the literature. Specifically, we examine the potential benefits of the subset quantile regression forecasts based on k-variate model forecasts ($k \geq 2$) under various combination methods (e.g. Mean, Median, Trimmed Mean, DMSFE, Cluster) relative to using subset linear regression forecasts based on k-variate models as proposed in EGT or relative to several combination methods of univariate linear and/or quantile models as proposed in MPVV.

4.2 Performance of Complete Subset Linear Regression Models

First, we discuss the out-of-sample performance of the forecasts obtained by subset linear regressions under various combination schemes. Table 1 presents the R^2_{OS} statistics of all subset regressions relative to the historical average benchmark model for the out-of-sample period 1965:1-2010:4. Positive values of R^2_{OS} indicate superior forecasting performance of the predictive models with respect to the historical average forecast. The statistical significance of the corresponding forecasts is assessed by using the Clark and West (2007) MSFE-adjusted statistic. The second column of the Table reports the R^2_{OS} generated by simply averaging the forecasts (Mean combination method) produced by subset linear regressions for various values of k. This experiment coincides with the framework of EGT and suggests that the subset linear regression with $k = 2$ generates the largest R^2_{OS} value (4.10%). Similarly to EGT, subset regression forecasts with $k \leq 6$ produce positive R^2_{OS} values, while the out-of-sample forecasting ability of subsets deteriorates markedly for $k \geq 7$.

[TABLE 1 AROUND HERE]

Next, we focus on alternative (to the Mean) combination methods such as the Median, Trimmed Mean, DMSFE and the Cluster combining schemes within the subset linear regression

[12] A brief description of the Clark and West (2007) test is given in Appendix B. Note that the critical values of the CW test are approximate in our setting, since the models we intend to compare are, in general, not nested. Alternatively, Bootstrap critical values can be employed; however, our recursive out-of-sample experiment is computationally very demanding.

approach. Overall, the largest R^2_{OS} values occur for the $k = 2$ subset, with the exception of the Cluster schemes where the largest R^2_{OS} occur for $k = 1$. For these subsets ($k = 2$ or $k = 1$), most of the combining methods produce statistically significant positive R^2_{OS} values, while four of them, namely the Median, Trimmed Mean, DMSFE(0.9) and DMSFE(0.5) provide higher values of R^2_{OS} than that of the best ($k = 2$) subset regression based on the Mean combination scheme. A comparison of the different combination techniques suggests that the DMSFE(0.5) scheme, which penalizes more recent forecasting accuracy, ranks first followed by the Median combination scheme. These methods provide the highest R^2_{OS} values of 4.58% and 4.40%, respectively. The results of Table 1 in general indicate that employing alternative weighting schemes can lead to improved forecasting performance relative to simple averaging.

4.3 Performance of Complete Subset Quantile Regression Models

In this subsection, we evaluate the forecasting performance of the proposed subset quantile regression models based on the RFC and the QFC approach.

4.3.1 Robust Forecast Combination approach

Our RFC approach employs either a fixed weighting (FW) or a time-varying weighting (TVW) scheme to construct robust point forecasts from each subset quantile regression. Then, these robust forecasts are combined into final point forecasts by employing the combination schemes outlined in Appendix A.1. Table 2 reports the R^2_{OS} statistics and the respective p-values of the Clark and West (2007) test for the subset quantile regression models based on the RFC approach for the three fixed weighting schemes, i.e. the FW1 scheme (Panel A), the FW2 scheme (Panel B) and the FW3 scheme (Panel C). Results are reported for various combination methods, namely the Mean, Median, Trimmed Mean, DMSFE and Cluster, based on $k = 1$ to $k = 12$ subset quantile regression forecasts.

Several findings emerge from this analysis. First, we observe that combining the forecasts of a subset of $k = 3$ quantile regression models produces higher R^2_{OS} values for almost all the combination methods with the exception of the Cluster(3) method for all fixed weighting schemes. Second, for this subset, i.e. $k = 3$, all the combination methods based on the robust quantile regression models generate higher R^2_{OS} values than the corresponding combining methods based on the best $k = 2$ subset linear regression models, indicating the superior forecasting ability of the proposed RFC quantile approach. Third, a comparison of the different combination methods suggest that the Median combination technique outperforms the alternative combination methods for the FW1 and FW2 schemes, generating R^2_{OS} values of 5.06% and 4.90%, respectively, while the Trimmed Mean combination method provides the highest R^2_{OS} statistic of 5.12% and outperforms the competing combination methods for the third fixed weighting scheme.

[TABLE 2 AROUND HERE]

Next, we present the out-of-sample performance of the subset quantile regression forecasts based on the time-varying weighting schemes TVW1-TVW3 (Table 3, Panels A-C). Three combination methods can be used in the time-varying weighting framework; the Mean, Median and the Trimmed Mean. Based on the results of Table 3, we observe that the largest R^2_{OS} values occur for $k = 2$ or $k = 3$ subsets. For these subsets ($k = 2$ and $k = 3$), all the combining methods, i.e. the Mean, Median and Trimmed Mean, generate statistically significant positive R^2_{OS} values, which are higher than the corresponding R^2_{OS} values of the combining methods based on the best ($k = 2$) subset linear regression model (see Table 1). For these best subsets, the Median and the Trimmed Mean combination methods seem to outperform the Mean combination scheme since they produce higher R^2_{OS} values. The most striking result is the R^2_{OS} statistic of 5.59% obtained by the Median combination of forecasts of the $k = 2$ subset quantile regression models under the TVW1 scheme. Overall, our findings indicate superior predictive ability of the RFC approach on the basis of the first time-varying weighting scheme.

[TABLE 3 AROUND HERE]

4.3.2 Quantile Forecast Combination approach

We turn our attention to the results of the subset quantile regression models based on the QFC approach. According to this approach, the quantile forecasts obtained from different k-variate predictive model specifications are first combined employing several combination schemes. These schemes are either simple methods such as the Mean, Median and Trimmed Mean, or are based on the asymmetric linear loss function such as the DALFE and the AL Cluster methods. Then, robust point forecasts are obtained by synthesizing the different quantile forecasts employing either fixed weighting schemes or time-varying weighting schemes, thus exploiting the entire distributional information.

Table 4 reports the out-of-sample performance of the subset quantile regression forecasts obtained by the QFC approach using fixed weighting schemes (FW1-FW3). The results of Table 4 (Panel A - Panel C) indicate that high positive R^2_{OS} values are obtained by using $k = 2$, $k = 3$ and $k = 4$ subsets for all weighting schemes FW1-FW3. In particular, for $k = 3$ subsets almost all of the combining methods (except for the Cluster method) produce the highest positive R^2_{OS} values, which are larger than those of the best ($k = 2$) subset linear regression model (see Table 1) and similar or even higher than the corresponding R^2_{OS} values of the best ($k = 3$) subset quantile regression forecasts based on the RFC approach (see Table 2). Among the various combination methods, the Median combination scheme ranks first, since, for the best $k = 3$ subset, generates the highest R^2_{OS} values ranging from 5.22% for FW1 and FW2 to 5.32% for FW3 scheme. Second ranks the DALFE(0.5) method which produces R^2_{OS} values ranging from 4.93% for FW2 to 5.22% for FW3 scheme.

[TABLE 4 AROUND HERE]

Finally, Table 5 (Panels A-C) presents the results obtained by the subset QFC approach using time-varying weighting schemes (TVW1-TVW3). Three combination methods, namely the Mean, Median and Trimmed Mean, are used in this approach. Based on the results of Table 5, we observe that the subset quantile regression forecasts with $k = 2$ for QFC-TVW1 and QFC-TVW2 and with $k = 2$ or $k = 3$ for QFC-TVW3 generate statistically significant positive R^2_{OS} values. For these subsets ($k = 2$ or $k = 3$), the Median combination method outperforms the Mean and the Trimmed Mean combination schemes since it generates higher R^2_{OS} values. More importantly, the QFC-TVW1 approach based on the Median combination of $k = 2$ subsets of predictors produce the highest R^2_{OS} of 5.71% among the different forecasting approaches considered in our analysis (see Table 5, Panel A). These findings suggest that more promising results, i.e. best out-of-sample performance, are obtained by applying the proposed subset quantile regression models based on the QFC approach under the Median combination method for the first time-varying weighting scheme.

[TABLE 5 AROUND HERE]

5 Real time Selection of k

Our empirical findings (Section 4) suggest that the predictive performance of our subset quantile regression approach depends on the choice of the value of k. Therefore, it is important to develop a real time algorithm of selecting k recursively, based on the past history of excess returns and predictive variables, in order to produce 'optimal forecasts'. Since our proposed methodology involves forecasting an array of quantiles, it is quite interesting to examine whether the selected value of k varies across quantiles of returns, thus revealing a further source of information that can be exploited within our proposed framework. Our algorithm is flexible enough to allow for variability of the selected k across quantiles and, therefore, information on the best complete subset for each quantile of the return distribution can be incorporated within our approach.

5.1 Algorithm for selecting k

In this subsection we propose a likelihood-based (Bayesian) method for selecting k in real time. The experiment we conduct is naturally designed in the context of our QFC forecasting approach. At each time point in the out-of-sample period, indexed by $t + 1$, we compute the posterior probabilities of all values of k ($k \in \{1, 2, ..., K\}$), based on the data up to time t, for a set of quantiles. Then, for each quantile, τ, we select the most probable value of k and produce a quantile forecast at time $t + 1$, $\hat{r}_{t+1}(\tau)$, based on the selected complete subset. These quantile forecasts are then combined according to the fixed weighting and time-varying weighting schemes of Section 3 in order to produce 'optimal' QFC forecasts in real time.

Under the Bayesian approach to inference, uncertainty about any quantity of interest is represented by probability distributions. In regression variable selection problems there is uncertainty about the model specification. In our setting, it is of particular interest to quantify the uncertainty about the complete subset that will be used for predicting each quantile of returns. Therefore, in a Bayesian context, the random quantities of interest are the model specification, representing the set of predictors included in the jth model and denoted by m_j, $j = 1, ..., M$, $M = \sum_{i=1}^{K} n_{i,K}$, the value of k and the totality of the model parameters associated with the τth quantile regression, denoted by $\theta^{(\tau)}$. After specifying appropriate prior distributions for these quantities, $P(m_j)$, $P(k|m_j)$ and $f(\theta^{(\tau)}|m_j,k)$, their joint posterior distribution is given by

$$f(m_j,k,\theta^{(\tau)}|r_{1:t}) \propto P(m_j)P(k|m_j)f(\theta^{(\tau)}|m_j,k)L^{(\tau)}(r_{1:t}|m_j,k,\theta^{(\tau)}),$$

where $L^{(\tau)}(r_{1:t}|m_j,k,\theta^{(\tau)})$ is the likelihood of the data up to time t under the τth quantile regression (Equation 5), based on the asymmetric Laplace density (4). Dependence on the set of predictors has been suppressed for simplicity. Then, the marginal posterior distribution of k, under the τth quantile regression, is obtained as

$$P^{(\tau)}(k|r_{1:t}) \propto \sum_{j=1}^{M} P(m_j)P(k|m_j)\int f(\theta^{(\tau)}|m_j,k)L^{(\tau)}(r_{1:t}|m_j,k,\theta^{(\tau)})d\theta^{(\tau)}.$$

The integral $\int f(\theta^{(\tau)}|m_j,k)L^{(\tau)}(r_{1:t}|m_j,k,\theta^{(\tau)})d\theta^{(\tau)}$ is the marginal likelihood of the data under the τth quantile regression with k predictors and model specification m_j, i.e. $L^{(\tau)}(r_{1:t}|m_j,k)$. In this paper, we estimate the marginal likelihood by the BIC approximation which is given by

$$\widehat{L}^{(\tau)}(r_{1:t}|m_j,k) = \exp\{L^{(\tau)}(r_{1:t}|m_j,k,\widehat{\theta}^{(\tau)}) - k\ln(t)/2\},$$

where $\widehat{\theta}^{(\tau)}$ denotes the ML estimate of $\theta^{(\tau)}$, obtained as discussed in Subsection 2.2. Alternatively, the marginal likelihood of quantile regression models can be estimated by Laplace approximation (see Meligkotsidou, Vrontos and Vrontos, 2009).

The prior specification we consider is the following. The prior probability of the jth model is taken to be $P(m_j) = \pi^{k_j}(1-\pi)^{K-k_j}$, where π is the prior probability of including a predictor in the model, which is taken fixed and prespecified, and k_j is the number of predictors included in model m_j. In our analysis we consider two values of π. First, we set π equal to 1/2, thus reflecting complete prior ignorance about the model specification. Second, to slightly penalize models with too many predictors, we set π equal to 1/3, since it is known from previous studies (see Goyal and Welch, 2008, and Elliott, Gargano and Timmermann, 2013) that the best performing forecasts arise from models including fewer predictors. The prior probability of k given the model specification m_j is then $P(k|m_j) = 1$, if $k_j = k$, and $P(k|m_j) = 0$, otherwise. This prior structure leads to the joint prior of k, m_j being $P(k,m_j) = \pi^{k_j}(1-\pi)^{K-k_j}I(k_j = k)$ and to the natural Binomial(K,π) marginal prior on k. Then, the marginal posterior distribution of

k, under the τth quantile regression is given by

$$P^{(\tau)}(k|r_{1:t}) \propto \pi^k (1-\pi)^{K-k} \sum_{j=1}^{M} \widehat{L}^{(\tau)}(r_{1:t}|m_j, k) I(k_j = k).$$

Below we present and discuss the results of our likelihood-based approach to selecting k for the fixed and time-varying weighting schemes of Section 3 and the respective combining methods (see Appendix A.2).

5.2 Algorithm Performance

To gain some insight on the selected values of k for the quantiles of interest, Figures 1 and 2 plot the selected values under the above specified prior distribution with π=1/2 and π=1/3, respectively. It is evident that, at each time-point in both the holdout and the out-of-sample period, the selected value of k varies across quantiles. In general, larger values of k are selected for the extreme quantiles (τ=0.10 and τ=0.90). For the remaining ones, our algorithm almost always chooses $k = 2, 3, 4$ in the out of sample period. Thus, the produced 'optimal' forecasts are based on the combination of quantile forecasts obtained from k different complete subsets for all quantiles considered in our analysis. Moreover, some large values of k are selected for the quantiles of the left part of the return distribution (τ=0.10, 0.25, 0.33) in the holdout period. This may be due to the weaker likelihood information (i.e. likelihoods formed on smaller samples) used for selecting k throughout the holdout period. Finally, it is interesting to note that the selected values of k are slightly lower if the prior probability of inclusion is set to 1/3, thus penalizing the larger values of k.

[FIGURE 1 AROUND HERE]

[FIGURE 2 AROUND HERE]

Table 6 reports the out-of-sample performance of the 'optimal' QFC forecasts based both on fixed weighting schemes (FW1-FW3) and time-varying weights (TVW1-TVW3), under both prior specifications considered (i.e. π=1/2 (Panel A) and π=1/3 (Panel B)). The results of Table 6 reveal that our likelihood-based approach to selecting k in real time is extremely successful, since the values of R^2_{OS} obtained under all weighting schemes and for all combining methods are very high. Regarding the fixed weighting schemes, the largest R^2_{OS} values are obtained for the

Median combining method, being in all cases close to or higher than 4%, with the highest value being equal to 4.58% (for the FW2 scheme, under π=1/2). In accordance with our statistical significance results (Table 4), the DALFE(0.5) method ranks second with R^2_{OS} values very close to those obtained by the Median combining method. It is interesting to note that the results of the recursive k-selection exercise are quite robust across the combining methods considered, apart from the AL Cluster(3) method. Moreover, it appears that the FW2 scheme constantly outperforms the other two schemes of producing robust point forecasts based on fixed weights.

Similar findings pertain with respect to our TVW forecasts.[13] More in detail, the largest R^2_{OS} values are obtained for the Median combining method, ranging from 3.34% (for the TVW1 scheme, under π=1/2) to 4.33% (for the TVW2 scheme, under π=1/3), while the TVW2 scheme constantly outperforms the other two time-varying weighting schemes. In the time-varying weights framework, though, the results are slightly better in the case that the prior probability of inclusion is set to 1/3. This may be attributed to the fact that some very large values of k are selected throughout the holdout period, possibly due to weak likelihood information, especially in the case of π=1/2.

[TABLE 6 AROUND HERE]

In conclusion, let us note that the findings of our recursive experiment are very encouraging, since they show that the proposed approach of selecting k in real time, based only on the past history of the data, produces particularly well-performing forecasts and that these results are very robust to the choice of weighting scheme and combining method.

6 Economic Evaluation

Campbell and Thompson (2008) and Rapach, Strauss and Zhou (2010) suggest that even small predictability gains, in a statistical sense, can give an economically meaningful degree of return predictability providing increased portfolio returns for a mean-variance investor that maximizes expected utility. We follow this utility-based approach within this stylized asset allocation framework in order to rank the performance of competing models in a way that captures the

[13] Recall that only the Mean, Median and Trimmed Mean combining methods can be used within the time-varying weighting framework which requires a holdout period for the construction of robust point forecasts.

risk return trade-off.[14] Moreover, we not only exploit the information content in our forecasts for the expected equity premium but for the expected volatility of returns, as well. This is done by constructing robust volatility forecasts via a set of quantile forecasts utilizing the selection algorithm introduced in Section 5. Below, we outline our framework for measuring economic value along with the proposed framework for volatility forecast construction.

6.1 The framework for measuring economic value

Consider a risk-averse investor who constructs a dynamically rebalanced portfolio consisting of the risk-free asset and one risky asset. Her portfolio choice problem is how to allocate wealth between the safe (risk-free Treasury Bill) and the risky asset (stock market), while risk stems from the uncertainty over the future path of the stock market (both in terms of future returns and the uncertainty surrounding them). This approach involves only one risky asset and as such it can be thought of as a standard exercise of market timing in the stock market. In a mean-variance framework, the solution to the maximization problem of the investor yields the following weight (w_t) on the risky asset

$$w_t = \frac{E_t(r_{t+1})}{\gamma Var_t(r_{t+1})} = \frac{\widehat{r}_{t+1}}{\gamma Var_t(r_{t+1})}, \tag{6}$$

where E_t and Var_t denote the conditional expectation and variance operators, r_{t+1} is the equity premium and γ is the Relative Risk Aversion (RRA) coefficient that controls the investor's appetite for risk (Campbell and Viceira, 2002; Campbell and Thompson, 2008; Rapach, Strauss and Zhou, 2010). The conditional expectation $E_t(r_{t+1})$ of each model is given by the 'optimal' forecast from the specific model, $\widehat{r}_{t+1}$, and the variance, $Var_t(r_{t+1})$ is calculated using four alternative ways. The first method we employ is the ten-year rolling window of quarterly returns ($\widehat{\sigma}^2_{1,t+1}$). The remaining volatility forecasts are constructed using the interval approximation approach of Pearson and Tukey (1965). Specifically, we employ the following approximations

[14] This utility-based approach, initiated by West et al. (1993), has been extensively employed in the literature (Fleming, Kirby and Ostdiek, 2001; Marquering and Verbeek, 2004; Della Corte, Sarno and Tsiakas, 2009; Della Corte, Sarno and Valente, 2010; Wachter and Warusawitharana, 2009).

to conditional standard deviation based on symmetrical quantiles as follows:

$$\widehat{\sigma}_{2,t+1} = \frac{\widehat{r}_{t+1}(0.99) - \widehat{r}_{t+1}(0.01)}{4.65}, \tag{7}$$

$$\widehat{\sigma}_{3,t+1} = \frac{\widehat{r}_{t+1}(0.975) - \widehat{r}_{t+1}(0.025)}{3.92}, \tag{8}$$

$$\widehat{\sigma}_{4,t+1} = \frac{\widehat{r}_{t+1}(0.95) - \widehat{r}_{t+1}(0.05)}{3.25}. \tag{9}$$

The denominators in the above formulae are based on the central distances between estimated quantiles under Pearson curves which are slightly different from a Gaussian curve. The forecasts for the quantiles of interest, τ=0.01, τ=0.025, τ=0.05, τ=0.95, τ=0.975 and τ=0.99, are based on the combination of quantile forecasts within the kth complete subset, with the values of k being optimally selected at each point of time employing our proposed selection algorithm. Figures 3 and 4 show the selected values of k for the six quantiles required for the calculation of the volatility forecasts under prior probability of inclusion equal to 1/2 and 1/3, respectively, over the out-of-sample period. It can be seen that the selected values of k for the right tail quantiles are, in general, larger than the respective values of k for the more central quantiles (see Figures 1 and 2). As previously, the values of k chosen by our algorithm are slightly lower if a smaller value for the prior probability of inclusion is considered.

[FIGURE 3 AROUND HERE]

[FIGURE 4 AROUND HERE]

Equation (6) implies that the optimal weights depend on both the conditional mean and variance and as a result on the respective forecasts each model/ specification gives. In this setting the optimally constructed portfolio gross return over the out-of-sample period, $R_{p,t+1}$, is equal to

$$R_{p,t+1} = w_t \cdot r_{t+1} + R_{f,t},$$

where $R_{f,t} = 1 + r_{f,t}$ denotes the gross return on the risk-free asset from period t to $t+1$.[15] Over the forecast evaluation period the investor with initial wealth of W_o realizes an average utility of

$$\overline{U} = \frac{W_o}{(P - P_0)} \left[\sum_{t=0}^{P-P_0-1} (R_{p,t+1}) - \frac{\gamma}{2} \sum_{t=0}^{P-P_0-1} (R_{p,t+1} - \overline{R}_p)^2 \right], \tag{10}$$

[15] We constrain the portfolio weight on the risky asset to lie between 0% and 150% each month, i.e. $0 \leq w_t \leq 1.5$.

where $R_{p,t+1}$ is the gross return on her portfolio at time $t+1$. At any point in time, the investor prefers the predictive model that yields the highest average realized utility.[16]

The economic value of our modeling approaches is assessed by comparing their average utility to the corresponding value obtained under the benchmark prevailing mean model. Our results are reported in the form of the annualized Certainly Equivalent Return (CER), i.e. the return that would leave an investor indifferent between using the prevailing mean forecasts versus the forecasts produced by one of our proposed approaches and is calculated as follows:

$$CER = \Delta\overline{U} = \overline{U^i} - \overline{U^{PM}}, \tag{11}$$

where $\overline{U^i}$ is the average realized utility over the out-of-sample period of any of our competing models/ specifications (i) and $\overline{U^{PM}}$ is the respective value for the prevailing mean (PM) model. If our proposed model does not contain any economic value, CER is negative, while positive values of the CER suggest superior predictive ability against the PM benchmark.

6.2 Empirical evidence on the economic value of predictive regressions

We assume that the investor dynamically rebalances her portfolio (updates the weights) quarterly over the out-of-sample period employing the forecasts given by the QFC approach and our selection algorithm for $\pi = 1/2$ and $\pi = 1/3$. Similarly to Section 4 and 5, the out-of-sample period of evaluation is 1965:1-2010:4 and the benchmark strategy against which we evaluate our forecasts is the PM model. For every model/specification we calculate the CER associated with each strategy calculated from Equation (11) setting RRA (γ) equal to 3. Table 7 reports our findings for the aforementioned prior specifications. Panels A-C and Panels D-F report CER in annualized percentage points for the fixed weighting schemes and the time-varying weighting schemes, respectively under the alternative variance forecasts. The columns labeled σ_1 refer to the rolling variance forecast, while σ_2 to σ_4 refer to the robust subset variance forecasts given by equations (7)-(9). CER_1 and CER_2 refer to the prior specifications of $\pi = 1/2$ and $\pi = 1/3$, respectively.

The most striking feature of Table 7 is the robustness of benefits generated to an investor willing to adopt our modelling approaches which range from 3.91% to an impressive 6.27%

[16] We standardize the investor problem by assuming $W_o = 1$.

per year. More in detail, the maximum CER is attained when the Median TVW1 scheme is employed in conjunction with a robust variance forecast given by (8) and a prior of $\pi = 1/3$, which penalizes large values of subsets. On the other hand, the minimum CER, albeit quite high, is attained under the FW scheme when the AL Cluster (3) FW1 scheme is employed combined with the rolling variance forecast and the same prior specification. Overall, the TVW schemes appear superior to their FW counterparts. The minimum benefits to an investor increase to 4.58% when TVW schemes are employed compared to 3.91% under FW specifications. When comparing the alternative prior specifications, the prior of $\pi = 1/3$ appears superior as it leads to greater gains in all the approaches considered with the exception of the FW and TVW3 schemes under a rolling variance specification scheme. In accordance with our findings from the statistical evaluation of the forecasts obtained under alternative combination methods, the DALFE(0.5) combining method emerges as the optimal one when FW schemes are considered, while the Median one generates the highest CERs among the TVW schemes. With respect to the alternative conditional variance specifications, we have to note that the proposed robust subset variance forecasts add significant economic value within our asset allocation framework. Further benefits are achieved when either σ_2 or σ_3 (given by equations (7)-(8)) are employed as opposed to σ_4 which employs closer to the central location quantile forecasts.

[TABLE 7 AROUND HERE]

7 Conclusions

In this study we propose a complete subset quantile regression approach to equity premium prediction. The aim of our analysis is to construct equity premium forecasts, which take into account the benefits emerging from the subset framework, the quantile regression framework and the information given by the potential predictors.

The quantile predictive approach proposed in this paper is based on the combination of the quantile forecasts, or the robust point forecasts, across complete subsets of model specifications that keep the number of predictors, k, fixed. Forecast combination is based on several well-established combining methods, while robust and accurate forecasts of the equity premium are constructed as weighted averages of a set of quantile forecasts by employing either fixed or time-varying weighting schemes.

An important contribution of this study is the development of a likelihood-based method for selecting the value of k recursively. The proposed algorithm is able to identify the best subset for predicting each quantile of the return distribution in real time, based only on the past history of the data. Then, these 'optimal' quantile forecasts are combined to produce robust equity premium forecasts.

The results of our study are very promising. Our findings suggest that our complete subset quantile regression framework achieves superior predictive performance relative to the historical average benchmark, the combination approach, and the subset linear regression approach, both in statistical and economic evaluation terms. More importantly, our economic evaluation results suggest that a mean-variance investor that adopts our framework can gain sizable benefits which range from 3.91% to 6.27% per year relative to a naive strategy based on the historical benchmark performance.

Appendix A. Forecast Combination Schemes

Combining individual models' forecasts can reduce uncertainty risk associated with a single predictive model and display superior predictive ability (Bates and Granger, 1969; Hendry and Clements, 2004). In Appendix A.1, we briefly discuss existing combination schemes that are appropriate for combining either subset mean regression forecasts or subset robust forecasts based on quantile regression models (RFC approach), while in Appendix A.2 we introduce the respective combining methods that are appropriate for producing combined subset quantile forecasts (QFC approach).

A.1. Combination Methods for Mean forecasting

The combination forecasts of r_{t+1}, denoted by $\hat{r}^{(C)}_{t+1}$, are weighted averages of the $k-$variate predictor individual forecasts within each subset, $\hat{r}_{i,t+1}$, $i = 1, \ldots, n_{k,K}$, of the form $\hat{r}^{(C)}_{t+1} = \sum_{i=1}^{n_{k,K}} w^{(C)}_{i,t} \hat{r}_{i,t+1}$, where $w^{(C)}_{i,t}, i = 1, ..., n_{k,K}$, are the *a priori* combining weights at time t for each specific subset, $k, k \leq K$.

The simplest combining scheme is the one that attaches equal weights to all k-variate models for a specific k, i.e. $w^{(C)}_{i,t} = 1/n_{k,K}$, for $i = 1, ..., n_{k,K}$, called the Mean combining scheme. The next schemes we employ are the Trimmed Mean and Median ones. The Trimmed Mean combination scheme sets $w^{(C)}_{i,t} = 0$ for the smallest and largest forecasts and $w^{(C)}_{i,t} = 1/(n_{k,K} - 2)$ for the remaining ones, while the Median combination scheme employs the median of the $\{\hat{r}_{i,t+1}\}_{i=1}^{n_{k,K}}$ forecasts.

The methods we describe below require a holdout out-of-sample period during which the combining weights are estimated. To this end, the first P_0 out-of-sample observations are employed as the initial holdout period over which we construct combination forecasts and the remaining $T - (T_0 + P_0) = P - P_0$ forecasts are available for evaluation. The second class of combining methods we consider, proposed by Stock and Watson (2004), suggests forming weights based on the historical performance of the individual models over the holdout out-of-sample period. Specifically, their Discount Mean Squared Forecast Error (DMSFE) combining method suggests forming weights as follows

$$w^{(C)}_{i,t} = m^{-1}_{i,t} / \sum_{j=1}^{n_{k,K}} m^{-1}_{j,t},\ m_{i,t} = \sum_{s=T_0}^{t-1} \psi^{t-1-s} (r_{s+1} - \widehat{r}_{i,s+1})^2,$$

where ψ is a discount factor which attaches more weight on the recent forecasting accuracy of the individual models in the cases where $\psi \in (0,1)$. The values of ψ we consider are 1.0, 0.9 and 0.5. When ψ equals one, there is no discounting and the combination scheme coincides with the optimal combination forecast of Bates and Granger (1969) in the case of uncorrelated forecasts.

Finally, the third class of combining methods, namely the Cluster combining method, was introduced by Aiolfi and Timmermann (2006). In order to create the Cluster combining forecasts, we form L clusters of forecasts of equal size based on the MSFE performance. Each combination forecast is the average of the k-variate model forecasts in the best performing cluster. This procedure begins over the initial holdout out-of-sample period and goes through the end of the available out-of-sample period using a rolling window. In our analysis, we consider $L = 2, 3$.

A.2. Combination Methods for Quantile Forecasting

The DMSFE, Cluster and Principal Components combining methods have been designed in the framework of standard linear regression, in order to construct forecasts that exploit the entire set of predictive variables. The combining weights, $w_{i,t}^{(C)}$, are computed based on the MSFE, that is on a quadratic loss function that measures how close to the realized excess returns the individual forecasts are. These methods are appropriate within the framework of the RFC approach since, according to this approach, several robust point forecasts are first obtained from different single predictor quantile regressions and then these point forecasts are combined in order to exploit information from the available set of predictors. However, these combining schemes are not appropriate for combining predictor information within the QFC approach since variable information is now combined in the context of forecasting several quantiles of returns rather than producing point forecasts. In this case, the MSFE is no longer suitable for measuring the performance of the produced forecasts and has to be replaced by a metric based on the asymmetric linear loss function.

Below we describe how we modify the existing combining methods in order to produce quantile forecasts that exploit variable information. The combined quantile forecasts, $\hat{r}_{t+1}^{(C)}(\tau)$, are weighted averages of the form $\hat{r}_{t+1}^{(C)}(\tau) = \sum_{i=1}^{n_{k,K}} w_{i,t}^{(C)} \hat{r}_{i,t+1}(\tau)$, where the combining weights, $w_{i,t}^{(C)}$, have to be computed based on the check function (3).

First, we introduce the Discount Asymmetric Loss Forecast Error (DALFE) combining method which suggests forming weights as follows

$$w_{i,t}^{(C)} = m_{i,t}^{-1} / \sum_{j=1}^{n_{k,K}} m_{j,t}^{-1}, \; m_{i,t} = \sum_{s=T_0}^{t-1} \psi^{t-1-s} \rho_\tau(r_{s+1} - \widehat{r}_{i,s+1}(\tau)),$$

where $\psi \in (0,1)$ is a discount factor. The combining weights are computed based on the historical performance of the individual quantile regression models over the holdout out-of-sample period and ψ is set equal to 0.5, 0.9 and 1.

We also modify the Cluster combining method by forming L clusters of forecasts based on their performance as measured by the asymmetric loss forecast error. The Asymmetric Loss Cluster (AL Cluster) combination forecast is the average of the individual quantile forecasts in the best performing cluster which contains the forecasts with the lower expected asymmetric loss values. We consider forming $L = 2, 3$ clusters.

Appendix B. The Clark and West (2007) test of equal forecasting ability.

Clark and West (2007) develop an adjusted version of the Diebold and Mariano (1995) and West (1996) statistic, namely the MSFE-adjusted statistic, which in conjunction with the standard normal distribution generates asymptotically valid inferences when comparing forecasts from nested linear models. Suppose that we want to evaluate the forecasts of a parsimonious model A relative to a larger model B. Under the null hypothesis of equal MSFE, model B should generate larger MSFE than model A, due to the estimation of additional parameters that introduces noise into the forecasts while these do not improve predictions. A smaller MSFE should not be considered as evidence of superiority of model A over B. In this respect, the testing procedure of Clark and West (2007) aims at correcting for the inflation in the MSFE of the larger model before evaluating the relative forecasting accuracy of the two models. Let $\hat{r}_{A,t+1}$ and $\hat{r}_{B,t+1}$ denote the one-step ahead forecasts for r_t obtained from models A and B respectively. We define

$$f_{t+1} = (r_{t+1} - \hat{r}_{A,t+1})^2 - [(r_{t+1} - \hat{r}_{B,t+1})^2 - (\hat{r}_{A,t+1} - \hat{r}_{B,t+1})^2]$$

The test statistic of Clark and West, denoted as $MSFE - adjusted$, is given by the standard $t - statistic$ of the regression of $\{f_{s+1}\}_{s=T_0+P_0}^{T-1}$on a constant. Given that under the alternative hypothesis of the test, model B has lower MSFE than model A, this is an one-sided test. Clark and West (2007) recommend using 1.282, 1.645 and 2.326 as critical values for a 0.10, 0.05 and 0.01 test, respectively. Extensive simulations performed by them, which consider a variety of different processes and settings show that the aforementioned critical values provide reliable results.

References

[1] Aiolfi, M., and A. Timmermann.(2006) "Persistence in Forecasting Performance and Conditional Combination Strategies." Journal of Econometrics, 135, 31–53.

[2] Ang, A., and G. Bekaert. (2007) "Return Predictability: Is It There?." Review of Financial Studies, 20, 651–707.

[3] Baker, M., and J. Wurgler. (2000) "The Equity Share in New Issues and Aggregate Stock Returns." Journal of Finance, 55, 2219-2257.

[4] Bassett, W. G., and H-L. Chen. (2001) "Portfolio style: Return-based attribution using quantile regression." Empirical Economics, 26, 293-305.

[5] Bates, J. M., and C.W.J. Granger. (1969) "The combination of forecasts." Operational Research Quarterly, 20, 451–468.

[6] Baur, D. G.; T. Dimpfl; and R. Jung. (2012) "Stock return autocorrelations revisited: A quantile regression approach." Journal of Empirical Finance, 19(2), 254-265.

[7] Breiman, L. (1996)."Bagging predictors" Machine Learning, 36, 105–139.

[8] Buchinsky, M. (1994) "Changes in U.S. Wage Structure 1963-1987: An application of Quantile Regression." Econometrica, 62, 405-458.

[9] Buchinsky, M. (1995) "Quantile Regression Box-Cox Transformation model, and the U.S. wage structure, 1963-1987." Journal of Econometrics, 65, 109-154.

[10] Buchinsky, M. (1998), "Recent Advances in Quantile Regression Models: A Practical Guidline for Empirical Research." Journal of Human Resources, 33, 88-126.

[11] Campbell, J. Y. (1987) "Stock Returns and the Term Structure." Journal of Financial Economics, 18, 373–99.

[12] Campbell, J. Y., and J. H. Cochrane. (1999) "By force of habit: A consumption-based explanation of aggregate stock market behavior." The Journal of Political Economy, 107, 205–251.

[13] Campbell, J. Y., and R. J. Shiller. (1988) "Stock Prices, Earnings, and Expected Dividends." Journal of Finance, 43, 661–76.

[14] Campbell, J. Y., and R. J. Shiller. (1998) "Valuation Ratios and the Long-Run Stock Market Outlook." Journal of Portfolio Management, 24, 11–26.

[15] Campbell, J. Y., and S. B. Thompson.(2008) "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" Review of Financial Studies, 21, 1509–31.

[16] Campbell, J.Y., and L. Viceira. (2002) "Strategic Asset Allocation." Oxford University Press, Oxford.

[17] Campbell, J. Y., and T. Vuolteenaho. (2004) "Inflation Illusion and Stock Prices." American Economic Review, 94, 19–23.

[18] Cenesizoglu, T., and A. Timmermann. (2012) "Do Return Prediction Models Add Economic Value?" Journal of Banking and Finance, 36, 2974-2987.

[19] Chen, Q. and Y. Hong. (2010) "Predictability of equity returns over different time horizons: a nonparametric approach." Manuscript, Cornell University.

[20] Chuang, C.-C.; C.-M. Kuan; and H.-Y. Lin. (2009) "Causality in quantiles and dynamic stock return-volume relations." Journal of Banking and Finance, 33(7), 1351-60.

[21] Clark, T. E., and K. D. West.(2007) "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." Journal of Econometrics, 138, 291–311.

[22] Dangl, T., and M. Halling (2012), Predictive regressions with time-varying coefficients." Journal of Financial Economics, 106 (1), 157-181.

[23] Della Corte, P., L. Sarno, and I. Tsiakas (2009). "An Economic Evaluation of Empirical Exchange Rate Models." Review of Financial Studies, 22 (9) 3491-3530.

[24] Della Corte, P., L. Sarno, and G. Valente (2010) "A Century of Equity Premium Predictability and Consumption-Wealth Ratios: An International Perspective." Journal of Empirical Finance, 17, 313-331.

[25] Diebold, F. X., and R. S. Mariano.(1995), "Comparing Predictive Accuracy." Journal of Business and Economic Statistics, 13, 253–63.

[26] Engle, R. F., and S. Manganelli. (2004) "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles." Journal of Business and Economic Statistics, 22, 367-381.

[27] Elliott, G., A. Gargano, and A. Timmermann. (2013) "Complete Subset Regressions." Journal of Econometrics, 177(2), 357-373.

[28] Fama, E. F., and K. R. French. (1988) "Dividend Yields and Expected Stock Returns." Journal of Financial Economics, 22, 3–25.

[29] Fama, E. F., and K. R. French. (1989) "Business Conditions and Expected Returns on Stocks and Bonds." Journal of Financial Economics, 25, 23–49.

[30] Fama, E. F., and G. W. Schwert. (1977) "Asset Returns and Inflation." Journal of Financial Economics, 5, 115–46.

[31] Ferreira, M.I., and P. Santa-Clara. (2011) "Forecasting stock market returns: the sum of the parts is more than the whole." Journal of Financial Economics, 100, 514–537.

[32] Fleming, J.; C. Kirby; and B. Ostdiek.(2001), "The Economic Value of Volatility Timing." Journal of Finance, 56, 329-352.

[33] Gastwirth, J.L. (1966), "On robust procedures." Journal of the American Statistical Association, 61, 929-948.

[34] Goyal, A. and I. Welch.(2008) A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." Review of Financial Studies, 21, 1455–508.

[35] Granger, C. W. J., and R. Ramanathan. (1984) "Improved methods of combining forecasts." Journal of Forecasting, 3, 197-204.

[36] Guidolin, M.; S. Hyde; D. McMillan and S. Ono. (2009) "Non-linear predictability in stock and bond returns: When and where is it exploitable?" International Journal of Forecasting, 25, 373–399.

[37] Guidolin, M. and A. Timmermann. (2009) "Forecasts of US short-term interest rates: A flexible forecast combination approach." Journal of Econometrics, 150, 297–311.

[38] Hansen, B. (2008) "Least-squares forecast averaging." Journal of Econometrics, 146, 342-350.

[39] Hendry, D. F., and M. P. Clements. (2004) "Pooling of Forecasts." Econometrics Journal, 7, 1–31.

[40] Henkel, S.J.; J.S. Martin; and F. Nadari. (2011), "Time-varying short-horizon predictability." Journal of Financial Economics, 99, 560–580.

[41] Hoerl, A.E., Kennard, R.W. (1970) "Ridge regression: biased estimation for nonorthogonal problems" Technometrics 12, 55–67.

[42] Hsiao, C., and S. K. Wan. (2014) "Is there an optimal forecast combination?." Journal of Econometrics,178(2), 294-309.

[43] Judge, G. G, R. C. Hill, W. E. Griffiths, H. Lutkepohl; and T.-C. Lee.(1988) "Introduction to the Theory and Practice of Econometrics." New York, Wiley.

[44] Koenker, R., and G. Bassett (1978). "Regression Quantiles." Econometrica, 46, 33-50.

[45] Kothari, S., and J. Shanken. (1997) "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis." Journal of Financial Economics, 44, 169–203.

[46] Lettau, M., and S. C. Ludvigson. (2001) "Consumption, Aggregate Wealth, and Expected Stock Returns." Journal of Finance, 56, 815–49.

[47] Ludvigson, S. C., and S. Ng. (2007) "The empirical risk-return relation: a factor analysis approach." Journal of Financial Economics, 83, 171–222.

[48] Marquering, W., and M. Verbeek. (2004) "The Economic Value of Predicting Stock Index Returns and Volatility." Journal of Financial and Quantitative Analysis, 39, 407–29.

[49] Meligkotsidou, L., I. D. Vrontos, and S. D. Vrontos. (2009) "Quantile Regression Analysis of Hedge Fund Strategies." Journal of Empirical Finance, 16, 264-279.

[50] Meligkotsidou, L., E. Panopoulou, I. D. Vrontos, and S. D. Vrontos (2014). "A Quantile Regression Approach to Equity Premium Prediction", Journal of Forecasting, forthcoming.

[51] Menzly, L.; T. Santos; and P. Veronesi. (2004), Understanding Predictability." Journal of Political Economy, 112, 1 (2004), 1-47.

[52] Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou. (2014) "Forecasting the equity risk premium: the role of technical indicators." Management Science 60(7), 1772-91.

[53] Pearson, E. S. and J. W. Tukey (1965) "Approximate Means and Standard Deviations Based on Distances between Percentage Points of Frequency Curves." Biometrika 52, 533-546.

[54] Rapach, D., J. Strauss, and G. Zhou (2010) "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." Review of Financial Studies, 23, 2, 821-862.

[55] Rapach, D., and G. Zhou. (2012) "Forecasting stock returns." In preparation for G. Elliott and A. Timmermann, Eds., Handbook of Economic Forecasting, Vol. 2.

[56] Raftery, A., Madigan, D., Hoeting, J. (1997) "Bayesian Model Averaging for linear regression models". Journal of the American Statistical Association 97, 179–191.

[57] Spiegel, M.(2008) "Forecasting the equity premium: Where we stand today." The Review of Financial Studies, 21, 1453–1454.

[58] Stock, J. H., and M. W. Watson (2004) Combination Forecasts of Output Growth in a Seven-Country Data Set. Journal of Forecasting, 23, 405–30.

[59] Tibshirani, R. (1996) "Regression shrinkage and selection via the Lasso." Journal of the Royal Statistical Society. Series B 58, 267–288.

[60] Timmermann, A. (2006) "Forecast combinations." In Handbook of Economic Forecasting, Vol. I, G. Elliott, C. W. J. Granger and A. Timmermann, eds. Amsterdam, Elsevier.

[61] Tukey, J. W (1977) "Explanatory Data Analysis." Addison-Wesley, Reading, MA.

[62] Wachter, J., and M. Warusawitharana. (2009) "Predictable returns and asset allocation: Should a skeptical investor time the market?" Journal of Econometrics, 148 (2), 162-178.

[63] West, K. D. (1996) "Asymptotic Inference About Predictive Ability. Econometrica", 64, 1067–84.

[64] West, K., H. Edison and D. Cho (1993) "A Utility-based Comparison of Some Models of Exchange Rate Volatility." Journal of International Economics, 35, 23-46.

[65] Yu, K., and J. Zhang (2005). "A Three-Parameter Asymmetric Laplace Distribution and Its Extension." Communications in Statistics - Theory and Methods, 34, 1867-1879.

[66] Yu, K., Z. Lu, and J. Stander (2003), "Quantile regression: applications and current research areas." The Statistician, 52, 331-350.

[67] Yu, K., and R. A. Moyeed (2001), "Bayesian quantile regression." Statistics and Probability Letters, 54, 437-447.

Table 1. Out-of-sample performance of complete subset linear regression models

	Mean		*Median*		*Trimmed Mean*		$DMSFE(1)$		$DMSFE(0.9)$		$DMSFE(0.5)$		$Cluster(2)$		$Cluster(3)$	
k	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}
1	2.99	0.002	2.48	0.001	2.81	0.001	3.00	0.002	3.04	0.003	3.83	0.007	2.53	0.015	1.58	0.070
2	4.10	0.004	4.40	0.001	4.22	0.003	4.03	0.004	4.13	0.005	4.58	0.007	2.21	0.033	0.76	0.083
3	3.92	0.006	4.07	0.004	4.04	0.005	3.82	0.007	3.98	0.007	4.22	0.009	1.92	0.026	0.77	0.052
4	2.98	0.009	2.65	0.011	3.02	0.009	2.88	0.010	3.04	0.011	3.23	0.012	1.61	0.022	0.54	0.037
5	1.64	0.014	1.17	0.016	1.60	0.014	1.54	0.014	1.63	0.016	1.84	0.016	0.52	0.024	-0.69	0.040
6	0.07	0.020	-0.41	0.023	0.00	0.020	-0.02	0.020	-0.05	0.023	0.19	0.022	-1.17	0.031	-2.56	0.047
7	-1.70	0.027	-2.64	0.033	-1.78	0.027	-1.79	0.027	-1.93	0.031	-1.64	0.029	-3.15	0.039	-4.51	0.054
8	-3.72	0.035	-4.52	0.039	-3.83	0.036	-3.82	0.036	-4.05	0.041	-3.69	0.037	-5.58	0.049	-6.39	0.056
9	-6.10	0.046	-5.88	0.039	-6.29	0.046	-6.20	0.046	-6.47	0.051	-6.04	0.047	-8.17	0.058	-8.53	0.059
10	-8.98	0.058	-11.49	0.077	-9.28	0.058	-9.08	0.058	-9.32	0.062	-8.84	0.059	-10.98	0.068	-10.98	0.062
11	-12.53	0.072	-14.84	0.081	-13.06	0.071	-12.60	0.072	-12.75	0.074	-12.31	0.073	-13.92	0.073	-14.40	0.082
12	-16.95	0.090														

Notes: The Table reports the out-of-sample R^2 statistic with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2010:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 2. Out-of-sample performance of Robust Forecast Combination (RFC) approach-Fixed weighting (FW) schemes

	$Mean$		$Median$		$Trimmed\ Mean$		$DMSFE(1)$		$DMSFE(0.9)$		$DMSFE(0.5)$		$Cluster(2)$		$Cluster(3)$	
k	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}
Panel A: RFC-FW1																
1	2.01	0.021	1.51	0.039	1.76	0.030	2.11	0.018	2.20	0.019	3.26	0.013	2.49	0.018	1.01	0.128
2	4.22	0.003	4.51	0.001	4.36	0.003	4.23	0.003	4.29	0.004	4.68	0.007	3.26	0.019	2.22	0.042
3	4.83	0.003	5.06	0.002	4.98	0.003	4.77	0.003	4.85	0.004	4.88	0.005	3.46	0.011	2.72	0.018
4	4.35	0.003	4.30	0.004	4.45	0.003	4.26	0.004	4.40	0.004	4.34	0.005	3.41	0.007	2.94	0.010
5	3.30	0.005	3.25	0.005	3.31	0.004	3.20	0.005	3.34	0.005	3.25	0.006	2.84	0.006	2.01	0.011
6	1.93	0.006	1.66	0.006	1.90	0.006	1.84	0.006	1.97	0.007	1.87	0.007	1.67	0.007	0.89	0.011
7	0.49	0.007	0.00	0.008	0.46	0.007	0.40	0.008	0.49	0.008	0.42	0.008	0.24	0.009	-0.46	0.013
8	-1.17	0.009	-1.58	0.010	-1.22	0.009	-1.27	0.009	-1.21	0.010	-1.21	0.010	-1.28	0.010	-1.90	0.013
9	-3.35	0.012	-3.28	0.010	-3.47	0.012	-3.45	0.012	-3.40	0.014	-3.28	0.013	-3.61	0.013	-3.65	0.014
10	-6.28	0.018	-7.23	0.018	-6.60	0.018	-6.37	0.018	-6.28	0.019	-5.95	0.017	-6.55	0.018	-5.91	0.016
11	-10.52	0.030	-12.31	0.034	-11.06	0.029	-10.57	0.030	-10.41	0.030	-10.00	0.027	-11.45	0.031	-10.44	0.026
12	-15.47	0.041														
Panel B: RFC-FW2																
1	1.87	0.028	1.06	0.078	1.70	0.036	1.96	0.025	2.01	0.027	2.96	0.017	2.52	0.026	1.11	0.125
2	3.99	0.005	4.13	0.002	4.14	0.004	4.02	0.005	4.06	0.006	4.44	0.008	3.20	0.024	1.29	0.080
3	4.70	0.003	4.90	0.003	4.86	0.003	4.66	0.003	4.75	0.004	4.79	0.006	4.06	0.008	2.89	0.019
4	4.33	0.004	4.55	0.003	4.43	0.004	4.25	0.004	4.41	0.004	4.37	0.005	4.05	0.005	3.09	0.011
5	3.36	0.004	3.37	0.005	3.36	0.005	3.28	0.005	3.48	0.005	3.46	0.006	3.43	0.005	2.69	0.009
6	2.01	0.006	1.90	0.006	1.96	0.006	1.93	0.006	2.14	0.006	2.17	0.007	2.39	0.006	1.79	0.009
7	0.43	0.007	-0.02	0.008	0.38	0.007	0.35	0.007	0.55	0.008	0.65	0.008	0.83	0.007	0.35	0.010
8	-1.42	0.009	-1.72	0.010	-1.50	0.009	-1.50	0.009	-1.30	0.010	-1.09	0.009	-1.17	0.009	-1.22	0.010
9	-3.67	0.012	-3.54	0.010	-3.80	0.012	-3.76	0.012	-3.56	0.013	-3.21	0.012	-3.71	0.011	-3.10	0.010
10	-6.52	0.016	-7.65	0.015	-6.79	0.016	-6.60	0.016	-6.42	0.017	-5.90	0.015	-7.24	0.016	-5.64	0.011
11	-10.54	0.023	-12.99	0.028	-11.18	0.023	-10.57	0.024	-10.47	0.023	-9.94	0.022	-11.40	0.023	-10.56	0.020
12	-16.46	0.041														
Panel C: RFC-FW3																
1	2.29	0.010	1.65	0.025	1.98	0.016	2.37	0.009	2.44	0.011	3.42	0.010	2.47	0.020	1.79	0.073
2	4.42	0.002	4.65	0.001	4.55	0.002	4.41	0.003	4.44	0.003	4.83	0.006	3.05	0.022	1.84	0.057
3	4.96	0.003	5.08	0.002	5.12	0.002	4.88	0.003	4.94	0.003	5.01	0.005	3.25	0.012	2.14	0.026
4	4.43	0.003	4.43	0.004	4.53	0.003	4.32	0.004	4.43	0.004	4.43	0.005	3.11	0.009	2.63	0.013
5	3.34	0.005	3.15	0.005	3.35	0.005	3.24	0.005	3.34	0.006	3.32	0.006	2.54	0.008	1.61	0.014
6	1.95	0.007	1.73	0.007	1.92	0.007	1.86	0.007	1.94	0.008	1.94	0.008	1.36	0.010	0.39	0.015
7	0.51	0.009	-0.04	0.010	0.46	0.009	0.42	0.009	0.46	0.010	0.49	0.010	-0.11	0.011	-0.95	0.016
8	-1.13	0.011	-1.31	0.011	-1.18	0.011	-1.23	0.011	-1.21	0.012	-1.11	0.012	-1.65	0.013	-2.26	0.016
9	-3.18	0.014	-3.14	0.012	-3.29	0.014	-3.28	0.015	-3.29	0.016	-3.08	0.015	-3.72	0.016	-4.02	0.018
10	-5.94	0.020	-7.22	0.023	-6.18	0.020	-6.03	0.020	-6.00	0.021	-5.62	0.019	-6.74	0.023	-6.13	0.019
11	-9.86	0.031	-11.29	0.034	-10.25	0.031	-9.91	0.032	-9.82	0.032	-9.46	0.030	-10.54	0.031	-10.69	0.031
12	-14.69	0.042														

Notes: The Table reports the out-of-sample R^2 statistic of the Robust Forecast Combination (RFC) approach, under fixed weighting (FW) schemes with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2010:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 3. Out-of-sample performance of Robust Forecast Combination (RFC) approach-Time-varying weighting (TVW) schemes

	Panel A: RFC-TVW1						Panel B: RFC-TVW2						Panel C: RFC-TVW3					
	Mean		*Median*		*Trimmed Mean*		*Mean*		*Median*		*Trimmed Mean*		*Mean*		*Median*		*Trimmed Mean*	
k	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}
1	3.49	0.001	2.66	0.002	3.35	0.001	3.41	0.001	2.50	0.000	3.28	0.001	3.39	0.005	2.53	0.015	3.07	0.007
2	4.82	0.002	5.59	0.000	4.96	0.002	4.65	0.003	4.84	0.001	4.82	0.002	4.56	0.003	4.81	0.002	4.61	0.003
3	4.82	0.003	4.93	0.003	4.96	0.003	4.82	0.004	5.05	0.003	4.99	0.003	4.54	0.004	4.55	0.004	4.67	0.004
4	3.95	0.005	3.55	0.007	4.05	0.005	4.08	0.005	4.01	0.005	4.19	0.005	3.66	0.007	3.22	0.010	3.75	0.007
5	2.64	0.008	2.27	0.010	2.66	0.008	2.90	0.007	2.97	0.007	2.91	0.007	2.40	0.010	1.95	0.013	2.41	0.010
6	1.09	0.011	0.95	0.011	1.08	0.011	1.38	0.009	1.42	0.009	1.37	0.009	0.88	0.014	0.55	0.015	0.85	0.014
7	-0.46	0.014	-0.73	0.014	-0.46	0.014	-0.30	0.011	-0.68	0.012	-0.29	0.011	-0.68	0.017	-1.07	0.019	-0.70	0.017
8	-2.19	0.017	-2.50	0.017	-2.18	0.016	-2.23	0.014	-2.50	0.014	-2.23	0.014	-2.40	0.021	-2.63	0.021	-2.41	0.021
9	-4.37	0.022	-4.51	0.019	-4.43	0.021	-4.50	0.018	-3.92	0.013	-4.55	0.018	-4.55	0.026	-4.64	0.023	-4.63	0.026
10	-7.30	0.030	-8.28	0.030	-7.50	0.029	-7.39	0.023	-8.06	0.020	-7.57	0.023	-7.49	0.036	-8.14	0.035	-7.68	0.036
11	-11.64	0.049	-12.53	0.050	-11.97	0.046	-11.47	0.032	-13.59	0.036	-12.10	0.032	-11.70	0.057	-12.80	0.062	-11.96	0.054
12	-17.08	0.077					-17.97	0.063					-17.36	0.087				

Notes: The Table reports the out-of-sample R^2 statistic of the Robust Forecast Combination (RFC) approach, under time-varying weighting (TVW) schemes with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2010:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 4. Out-of-sample performance of Quantile Forecast Combination (QFC) approach-Fixed weighting (FW) schemes

	$Mean$		$Median$		$Trimmed\ Mean$		$DALFE(1)$		$DALFE(0.9)$		$DALFE(0.5)$		$AL\ Cluster(2)$		$AL\ Cluster(3)$	
k	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}
Panel A: QFC-FW1																
1	2.01	0.021	1.29	0.055	1.72	0.031	2.06	0.019	2.12	0.019	2.78	0.013	2.44	0.017	2.42	0.024
2	4.22	0.003	4.71	0.001	4.33	0.003	4.22	0.003	4.29	0.004	4.69	0.004	3.66	0.011	3.02	0.017
3	4.83	0.003	5.22	0.002	5.00	0.003	4.79	0.003	4.89	0.003	5.11	0.004	4.53	0.005	3.73	0.010
4	4.35	0.003	4.50	0.003	4.48	0.003	4.30	0.004	4.44	0.004	4.59	0.004	4.32	0.004	3.81	0.007
5	3.30	0.005	3.56	0.004	3.35	0.004	3.24	0.005	3.38	0.005	3.51	0.005	3.32	0.005	2.83	0.006
6	1.93	0.006	2.00	0.006	1.93	0.006	1.88	0.006	2.01	0.006	2.13	0.006	1.92	0.006	1.58	0.007
7	0.49	0.007	0.41	0.007	0.45	0.007	0.44	0.007	0.54	0.008	0.65	0.007	0.44	0.007	-0.09	0.009
8	-1.17	0.009	-1.34	0.009	-1.25	0.009	-1.23	0.009	-1.16	0.010	-1.07	0.009	-1.20	0.008	-1.58	0.009
9	-3.35	0.012	-3.40	0.011	-3.51	0.012	-3.40	0.012	-3.38	0.013	-3.31	0.012	-3.81	0.012	-3.66	0.011
10	-6.28	0.018	-6.95	0.016	-6.65	0.019	-6.33	0.018	-6.33	0.018	-6.20	0.017	-7.19	0.018	-6.74	0.017
11	-10.52	0.030	-12.77	0.037	-11.17	0.030	-10.56	0.030	-10.52	0.029	-10.31	0.028	-11.31	0.029	-11.05	0.025
12	-15.47	0.041														
Panel B: QFC-FW2																
1	1.87	0.028	0.71	0.124	1.58	0.042	1.90	0.027	1.96	0.027	2.56	0.018	2.29	0.023	2.87	0.024
2	3.99	0.005	4.54	0.002	4.10	0.004	3.99	0.005	4.06	0.005	4.44	0.006	3.47	0.014	3.00	0.020
3	4.70	0.003	5.22	0.002	4.87	0.003	4.67	0.003	4.78	0.004	4.93	0.004	4.36	0.005	4.01	0.008
4	4.33	0.004	4.48	0.004	4.46	0.003	4.29	0.004	4.43	0.004	4.49	0.004	4.48	0.004	3.93	0.005
5	3.36	0.004	3.60	0.004	3.42	0.004	3.32	0.004	3.49	0.004	3.55	0.005	3.71	0.004	3.32	0.004
6	2.01	0.006	2.06	0.006	2.01	0.006	1.97	0.006	2.15	0.006	2.23	0.005	2.55	0.004	2.16	0.005
7	0.43	0.007	0.24	0.008	0.44	0.007	0.39	0.007	0.57	0.007	0.70	0.007	0.98	0.005	0.45	0.005
8	-1.42	0.009	-1.52	0.009	-1.45	0.009	-1.47	0.009	-1.28	0.009	-1.09	0.008	-0.87	0.006	-1.15	0.006
9	-3.67	0.012	-3.36	0.009	-3.78	0.012	-3.72	0.012	-3.55	0.012	-3.27	0.011	-3.44	0.008	-3.04	0.007
10	-6.52	0.016	-7.41	0.015	-6.81	0.016	-6.57	0.016	-6.43	0.016	-6.07	0.014	-6.72	0.012	-5.84	0.009
11	-10.54	0.023	-12.53	0.026	-11.26	0.024	-10.55	0.023	-10.48	0.023	-10.18	0.021	-11.41	0.021	-11.32	0.020
12	-16.46	0.041														
Panel C: QFC-FW3																
1	2.29	0.010	1.50	0.032	1.97	0.015	2.34	0.010	2.37	0.010	3.02	0.008	2.47	0.015	2.40	0.022
2	4.42	0.002	4.80	0.001	4.52	0.002	4.41	0.002	4.46	0.003	4.85	0.004	3.79	0.009	3.08	0.015
3	4.96	0.003	5.32	0.002	5.14	0.002	4.92	0.003	5.00	0.003	5.22	0.003	4.50	0.005	3.67	0.010
4	4.43	0.003	4.58	0.003	4.56	0.003	4.38	0.004	4.48	0.004	4.60	0.004	4.22	0.005	3.69	0.007
5	3.34	0.005	3.56	0.005	3.40	0.005	3.29	0.005	3.38	0.005	3.45	0.005	3.16	0.006	2.67	0.008
6	1.95	0.007	1.97	0.007	1.94	0.007	1.91	0.007	1.97	0.007	2.02	0.007	1.74	0.007	1.34	0.009
7	0.51	0.009	0.38	0.009	0.46	0.009	0.47	0.009	0.50	0.009	0.52	0.009	0.25	0.009	-0.30	0.011
8	-1.13	0.011	-1.27	0.010	-1.21	0.011	-1.17	0.011	-1.18	0.012	-1.18	0.011	-1.34	0.010	-1.76	0.012
9	-3.18	0.014	-3.21	0.013	-3.32	0.015	-3.23	0.014	-3.28	0.015	-3.32	0.015	-3.84	0.014	-3.79	0.014
10	-5.94	0.020	-6.39	0.018	-6.25	0.021	-5.99	0.020	-6.05	0.021	-6.04	0.020	-7.07	0.021	-6.79	0.020
11	-9.86	0.031	-11.80	0.038	-10.40	0.031	-9.91	0.031	-9.92	0.032	-9.81	0.030	-10.97	0.033	-10.75	0.029
12	-14.69	0.042														

Notes: The Table reports the out-of-sample R^2 statistic of the Quantile Forecast Combination (QFC) approach under fixed weighting (FW) schemes with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2010:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 5. Out-of-sample performance of Quantile Forecast Combination (QFC) approach-Time-varying weighting (TVW) schemes

	Panel A: QFC-TVW1						Panel B: QFC-TVW2						Panel C: QFC-TVW3					
	Mean		*Median*		*Trimmed Mean*		*Mean*		*Median*		*Trimmed Mean*		*Mean*		*Median*		*Trimmed Mean*	
k	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}
1	3.93	0.001	3.41	0.001	3.64	0.001	3.85	0.001	2.99	0.000	3.54	0.001	3.15	0.010	2.73	0.010	2.87	0.011
2	5.06	0.002	5.71	0.001	5.23	0.002	4.99	0.002	5.54	0.001	5.18	0.002	4.45	0.004	5.02	0.002	4.63	0.003
3	4.61	0.004	4.95	0.003	4.84	0.004	4.96	0.003	5.33	0.002	5.17	0.003	4.46	0.005	4.76	0.004	4.65	0.004
4	3.73	0.006	3.88	0.006	3.86	0.006	4.16	0.005	4.34	0.005	4.32	0.005	3.67	0.007	3.65	0.007	3.74	0.006
5	2.35	0.009	2.55	0.008	2.39	0.009	2.80	0.007	3.05	0.007	2.87	0.007	2.26	0.010	2.44	0.010	2.26	0.010
6	0.78	0.013	0.99	0.011	0.78	0.013	1.28	0.010	1.44	0.009	1.33	0.010	0.57	0.015	0.72	0.014	0.55	0.015
7	-0.82	0.017	-0.64	0.014	-0.83	0.017	-0.40	0.013	-0.59	0.013	-0.35	0.012	-1.05	0.019	-1.03	0.017	-1.09	0.019
8	-2.49	0.020	-2.77	0.019	-2.57	0.020	-2.43	0.016	-2.50	0.015	-2.42	0.016	-2.84	0.025	-3.04	0.023	-2.92	0.025
9	-4.70	0.026	-4.89	0.023	-4.85	0.026	-4.81	0.021	-4.40	0.016	-4.89	0.021	-4.97	0.031	-5.18	0.030	-5.12	0.032
10	-7.75	0.036	-8.58	0.035	-8.12	0.037	-7.80	0.027	-8.54	0.024	-8.06	0.027	-7.84	0.041	-8.68	0.040	-8.20	0.043
11	-12.11	0.057	-14.50	0.070	-12.77	0.058	-11.88	0.037	-13.87	0.040	-12.65	0.037	-12.24	0.065	-14.41	0.077	-12.89	0.066
12	-17.08	0.077					-17.97	0.063					-17.36	0.087				

Notes: The Table reports the out-of-sample R^2 statistic of the Quantile Forecast Combination (QFC) approach under time-varying weighting (TVW) schemes with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2010:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 6. Out-of-sample performance of the 'optimal' QFC forecasts

Panel A: $\pi = 1/2$

	Mean		*Median*		*Trimmed Mean*		$DALFE(1)$		$DALFE(0.9)$		$DALFE(0.5)$		$AL\ Cluster(2)$		$AL\ Cluster(3)$	
	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}
FW1	3.97	0.005	4.26	0.004	4.11	0.005	3.92	0.005	4.03	0.005	4.22	0.005	3.79	0.007	3.11	0.011
FW2	4.31	0.004	4.58	0.003	4.42	0.003	4.28	0.004	4.42	0.004	4.47	0.004	4.53	0.003	4.12	0.005
FW3	4.01	0.005	4.29	0.004	4.15	0.005	3.96	0.005	4.03	0.006	4.20	0.006	3.64	0.008	2.94	0.013
TVW1	3.10	0.011	3.34	0.006	3.23	0.008										
TVW2	3.79	0.009	3.99	0.005	3.88	0.006										
TVW3	3.50	0.011	3.79	0.005	3.61	0.008										

Panel B: $\pi = 1/3$

	Mean		*Median*		*Trimmed Mean*		$DALFE(1)$		$DALFE(0.9)$		$DALFE(0.5)$		$AL\ Cluster(2)$		$AL\ Cluster(3)$	
	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}	R^2_{OS}	CW_{pv}
FW1	3.37	0.007	3.93	0.005	3.49	0.007	3.34	0.007	3.41	0.008	3.70	0.007	2.81	0.011	2.03	0.016
FW2	3.67	0.009	4.13	0.006	3.78	0.008	3.64	0.009	3.76	0.010	3.97	0.010	3.44	0.018	2.93	0.030
FW3	3.49	0.009	3.94	0.006	3.60	0.009	3.46	0.009	3.49	0.009	3.78	0.008	2.85	0.011	1.99	0.015
TVW1	3.10	0.015	3.71	0.009	3.28	0.013										
TVW2	3.85	0.008	4.33	0.006	3.97	0.008										
TVW3	3.18	0.012	3.64	0.009	3.33	0.011										

Notes: The Table reports the out-of-sample R^2 statistic of the Quantile Forecast Combination (QFC) approach under fixed (FW) and time-varying weighting (TVW) schemes with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2010:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 7. Economic evaluation of the 'optimal' QFC forecasts

	Panel A: FW1								Panel D: TVW1							
	σ_1		σ_2		σ_3		σ_4		σ_1		σ_2		σ_3		σ_4	
	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2
$Mean$	4.82	4.56	5.16	5.26	5.13	5.16	4.66	4.79	4.74	4.91	4.87	5.68	5.03	5.61	4.85	5.31
$Median$	4.95	4.92	5.23	5.51	5.38	5.39	4.81	5.09	4.91	5.43	5.13	6.27	5.50	6.20	5.15	5.86
$Trimmed\ Mean$	4.90	4.64	5.30	5.33	5.23	5.25	4.74	4.85	4.82	5.03	4.99	5.78	5.09	5.73	4.86	5.40
$DALFE(1)$	4.81	4.58	5.20	5.31	5.16	5.20	4.65	4.81								
$DALFE(0.9)$	5.05	4.83	5.42	5.35	5.33	5.34	4.90	4.96								
$DALFE(0.5)$	5.27	5.19	5.53	5.58	5.49	5.68	5.23	5.46								
$AL\ Cluster(2)$	4.83	4.28	5.30	5.13	5.29	5.13	5.32	5.27								
$AL\ Cluster(3)$	4.38	3.91	4.91	4.87	5.03	4.98	5.14	5.32								
	Panel B: FW2								Panel E: TVW2							
$Mean$	4.72	4.76	5.15	5.45	5.18	5.39	4.70	5.03	4.68	4.79	4.99	5.34	5.08	5.21	4.65	5.11
$Median$	4.89	5.33	5.25	5.79	5.43	5.85	4.84	5.50	4.75	5.16	5.14	5.87	5.28	5.71	4.66	5.48
$Trimmed\ Mean$	4.79	4.85	5.25	5.51	5.27	5.49	4.77	5.11	4.64	4.88	5.06	5.40	5.12	5.30	4.58	5.15
$DALFE(1)$	4.71	4.75	5.20	5.47	5.22	5.42	4.70	5.01								
$DALFE(0.9)$	5.01	4.98	5.48	5.56	5.43	5.52	5.00	5.29								
$DALFE(0.5)$	5.13	5.17	5.59	5.71	5.55	5.74	5.26	5.60								
$AL\ Cluster(2)$	4.99	4.55	5.48	5.49	5.53	5.48	5.66	5.76								
$AL\ Cluster(3)$	4.57	4.31	5.23	5.36	5.40	5.46	5.63	5.83								
	Panel C: FW3								Panel F: TVW3							
$Mean$	4.87	5.82	5.10	5.58	5.03	5.53	4.64	5.15	4.95	4.81	4.96	5.63	5.17	5.58	5.00	5.24
$Median$	4.96	5.89	5.20	5.68	5.32	5.81	4.90	5.52	5.15	5.33	5.11	6.23	5.54	6.14	5.24	5.76
$Trimmed\ Mean$	4.92	5.86	5.23	5.71	5.12	5.62	4.66	5.12	4.99	4.92	5.00	5.73	5.21	5.70	4.99	5.33
$DALFE(1)$	4.86	5.81	5.14	5.62	5.07	5.57	4.66	5.19								
$DALFE(0.9)$	5.06	5.99	5.31	5.79	5.20	5.70	4.77	5.19								
$DALFE(0.5)$	5.29	6.17	5.47	5.94	5.40	5.88	5.10	5.50								
$AL\ Cluster(2)$	4.65	5.48	5.06	5.54	5.07	5.55	4.93	5.34								
$AL\ Cluster(3)$	4.22	4.97	4.68	5.12	4.80	5.26	4.75	5.12								

Notes: CER denotes the Certainty Equivalent Return (reported in annualized percentage points) that an investor with mean-variance preferences and risk aversion coefficient of three would gain when employing the alternative specifications. CER_1 and CER_2 correspond to the selection of the best subset k on the basis of a prior of $\pi = 1/2$ and $\pi = 1/3$, respectively. The weight on stocks in the investor's portfolio is restricted to lie between zero and 1.5. .

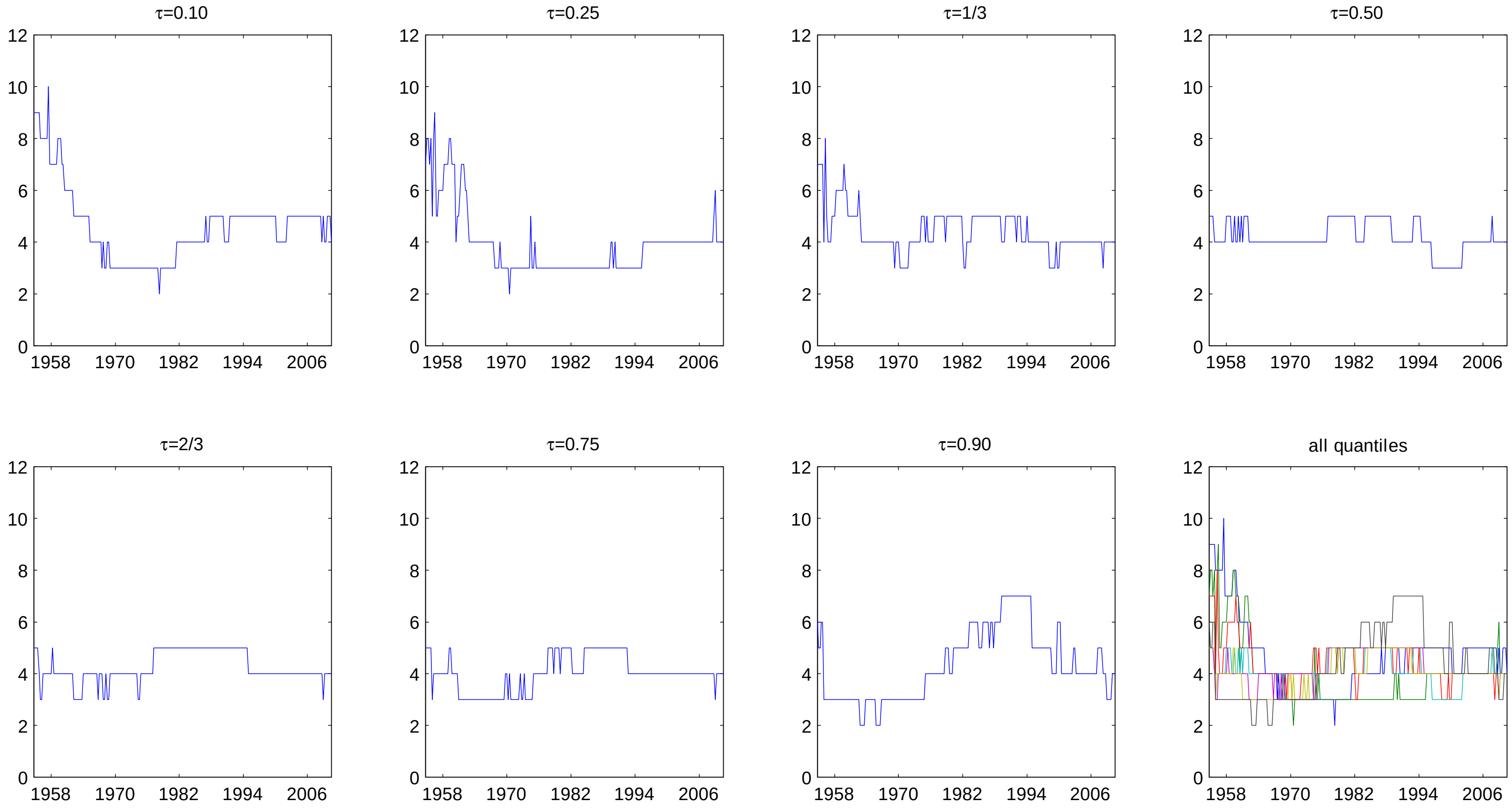

Figure 1: Selection of k over the holdout and out-of-sample period for the quantiles employed for equity premium prediction with prior probability of inclusion $\pi = 1/2$

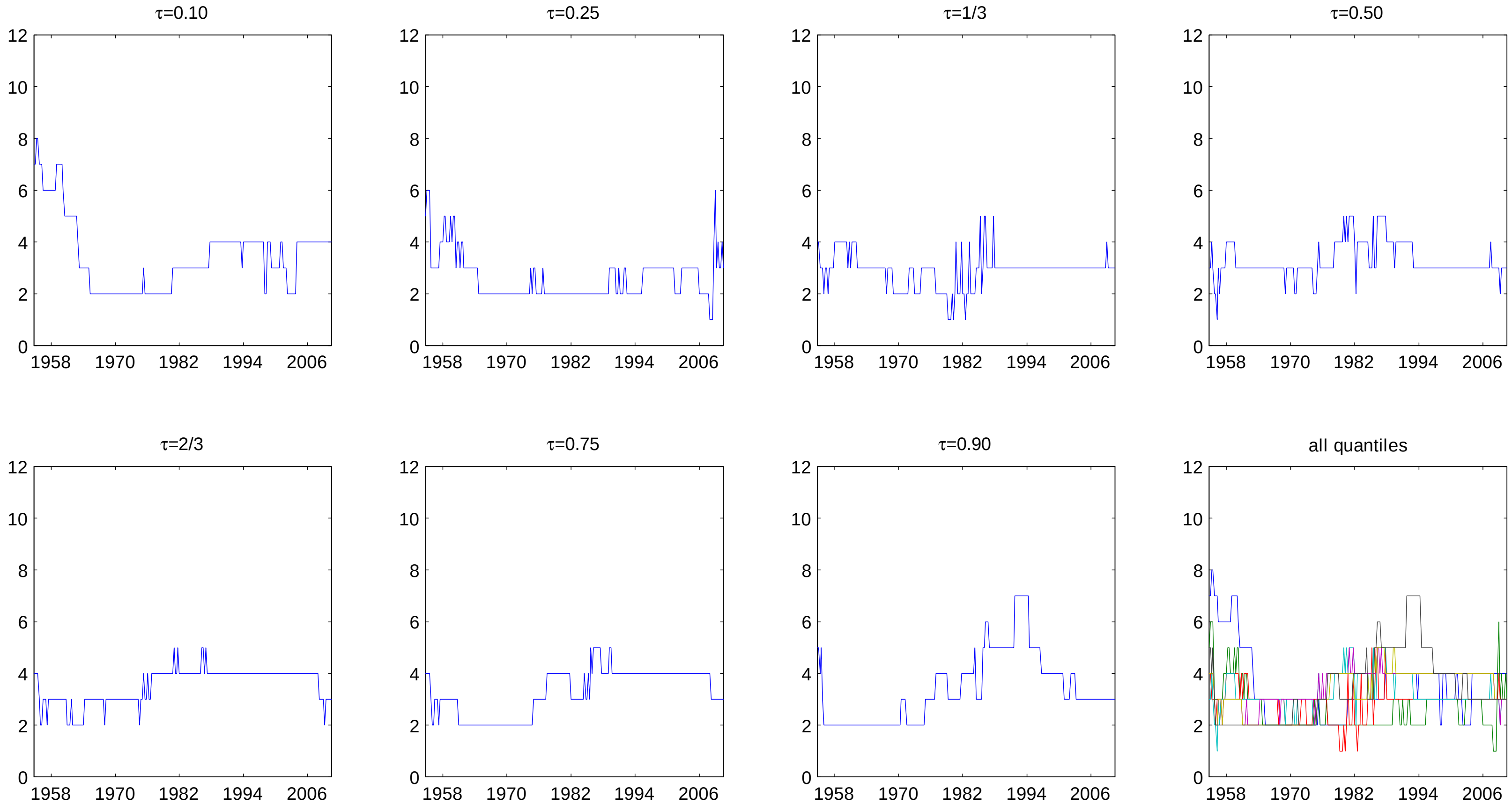

Figure 2: Selection of k over the holdout and out-of-sample period for the quantiles employed for equity premium prediction with prior probability of inclusion $\pi = 1/3$

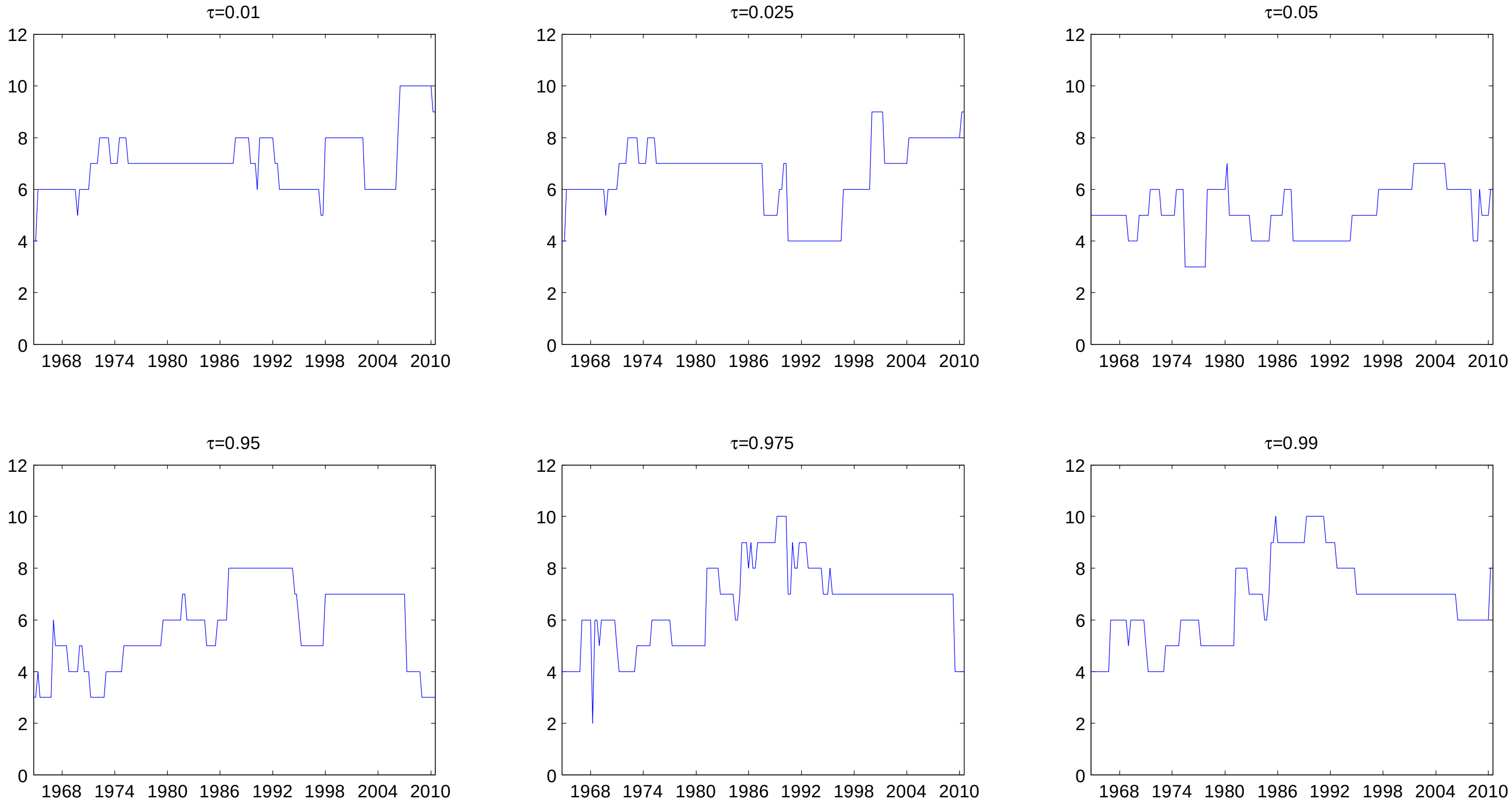

Figure 3: Selection of k over the out-of-sample period for the quantiles employed for calculating volatility forecasts with prior probability of inclusion $\pi = 1/2$

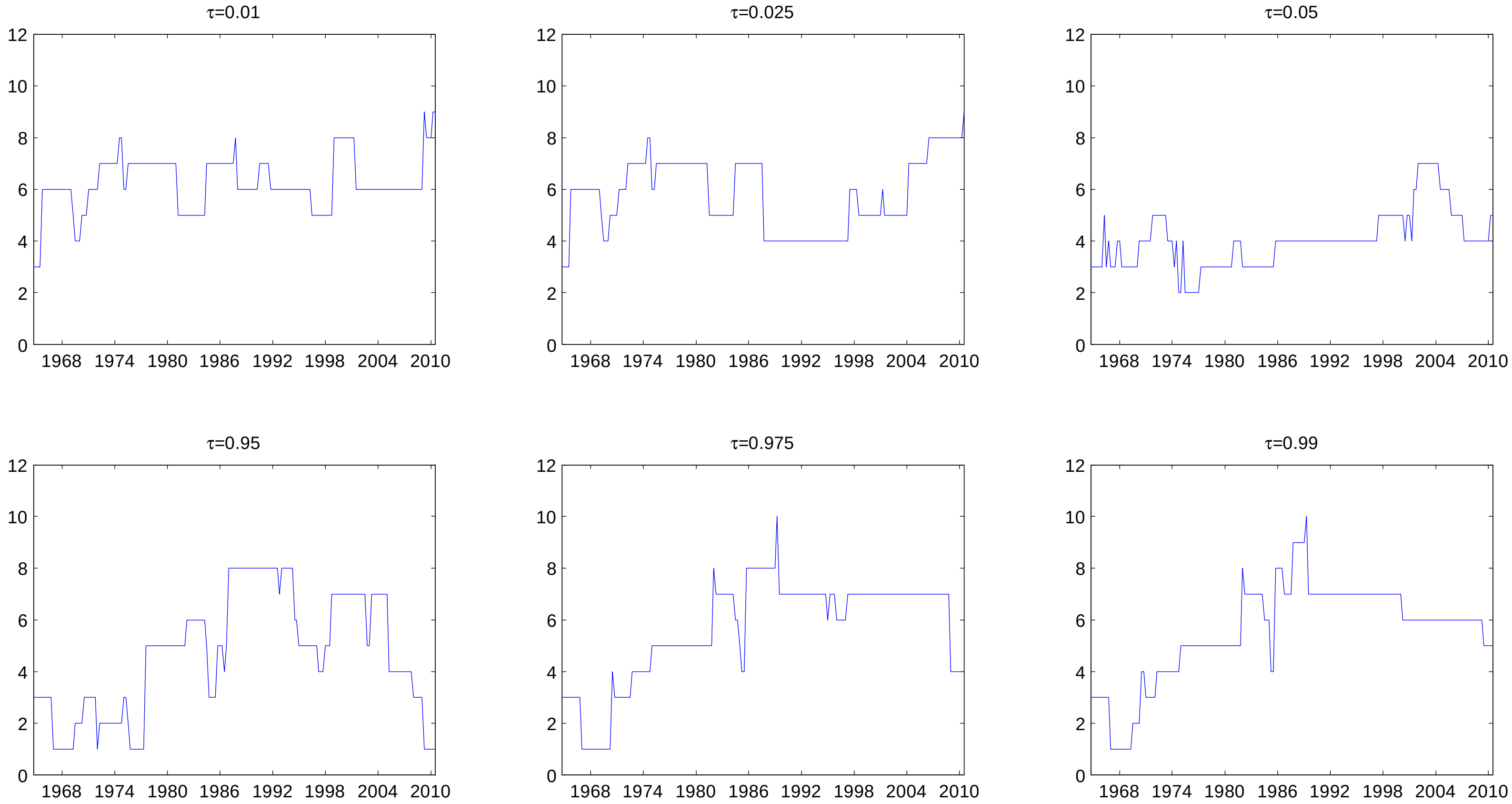

Figure 4: Selection of k over the out-of-sample period for the quantiles employed for calculating volatility forecasts with prior probability of inclusion $\pi = 1/3$